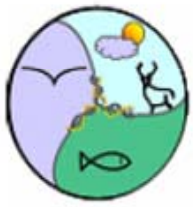
Fine-scale genetic structure of brown trout (*Salmo trutta*)

Annette Taugbøl

Master of Science thesis 2008



Department of Biology
University of Oslo

Centre for Ecological and Evolutionary Synthesis

Preface

This project is part of a long-term mark-recapture project at CEES: *Spatial and temporal population structure and breeding system in stream-living brown trout: an integrated ecological and genetic approach,* which started in 1997.

Firstly, I want to thank my supervisor Asbjørn Vøllestad for letting me take part in such an interesting study. Always with his door open, giving me guidance and answering all my questions patiently. I thank Dimitar Zerbezov, for letting me in on all the secrets in DNA amplification, PCR and genotyping, Dimitar, Johannes Holmen, Dorothee Ehrich, Kjartan Østby and Hege Gundersen for helping me and giving me valuable tips in different ways with the analysis. Similarly, Nanna Steen, Ave Tooming-Klunderud, Bård and Emelita Nerli for helping me, and training me use equipment in the lab.

I would also like to thank all my student friends, especially those who have and still are inhabiting study room 4419 ("Næsje et al"); Krisin Vesterkjær, Kirsten Kleveland, Inger Maren Riverud, Eirik Krogstad, Oda Bjerke, Truls Næsje, Gaute Thomassen, Evelyn Kristiansen and Maria Aasen for, when being there, always being in a good mood and up for discussions. Adine Thoresen, Gøril Hoel and Gunhild Borgersen for stopping by and sharing gossip and drama, Marthe Bråthen Lund for being drama and all the rest of the "bio-gang" for funny conversations sparkling up everyday life.

Gro, Lisa, Ingunn, Helene, Marianne, Siv, Silje, Torstein, Trond and Lars Helge of many that have socialized me outside school; thank you for making me laugh and take out frustration.

My dad, for proof-reading my English, soon sending me to a "was and were" course, for dinners and leftovers, My mum, Geir, Andrea, Karen, Temor, Mor and May for general support.

At last, but not least, I want to thank Kim Magnus Bærum (also a part of "Næsje et al.) for patience, support and helping me making the way more encouraging and joyful.

Annette Taugbøl,
University of Oslo, 2nd of January, 2008

Abstract

Many freshwater fishes are fragmented into local populations. Brown trout (*Salmo* trutta) is a fish that express high genetic heterogeneity and are known to be structured into genetically differentiated subpopulations over small limited areas. In this study, five sections in a river (Julussa) and three sections in a stream (Bellbekken) were sampled in order to investigate the genetic population structure. It was predicted that the two rivers would be genetically differentiated as they are separated by a small waterfall that could be a potential migration barrier hindering gene flow. It was also predicted that sample sites in Julussa, as this river had longer distances between sites, would be more differentiated than in Bellbekken, according to the isolation by distance model. Both predictions were investigated using F_{ST} values (fixation index) and two different assignment tests; one partially Bayesian and one fully Bayesian, in order to get information about migration between sample sites and population structure. The results confirmed a clear pattern of subdivision between the two rivers; the differentiation values between trout in Julussa and trout in Bellbekken were high (F_{ST} = 0.0216-0.0522), the migration was low and the structure test clearly divided the two rivers into two subpopulations. There was weak evidence for an isolation by distance genetic structure (r^2 = 0.069, p= 0.0790), indicating that other reasons than distance are more important for the pronounced subdivision. There was clear evidence for significant differentiation between sites in Bellbekken (F_{ST} = 0.0051 - 0.0213) and relatively high site specific assignments. This indicates genetic divergence in this stream, which is probably due to a low effective population size and low effective migration, resulting in genetic drift. There was no clear evidence for substructure within Julussa (F_{ST} = 0.0013 - 0.0100), which could indicate a larger effective population size and a higher rate of gene flow. Samples from the same site taken during sampling sessions were not significantly different, indicating a stable population structure, at least on a short time scale.

Table of contents

Introduction

The term “population” is widely used (Waples and Gaggiotti, 2006, and references therein) but definitions are equivocal. The expression is spanning from large scales as “a network of random mating individuals” (Wright, 1978) to finer scales where populations usually are identified as demes or subgroups, with varying degrees of relatedness due to different patterns of genetic and demographic connectivity (Sugg *et al.*, 1996). In a population, the group of randomly mating individuals is partly or totally physical reproductively isolated from other such groups within the same species. Different mating patterns (selection) and rates of gene flow (migration) between individuals within a population can split them into subpopulations (Wright, 1931). Depending on the sizes of the subpopulations and the amount of gene flow between them, genetic drift (random changes in allele frequencies) and mutations (Wright, 1931) can further contribute to heterogeneity between the local populations. High gene flow precludes local adaptation when it homogenizes the allele frequencies between subpopulations; it generates new polymorphism and increase local effective population size (opposing random genetic drift when gene flow is generating new gene combinations). Therefore, to maintain heterogenetic subpopulations, the amount of gene flow between them should be relatively low, when, on average, one individual exchanged every generation between local populations will prevent genetic drift when only drift is operating (Slatkin, 1985). However, genetic subdivision in form of statistically significant allele frequency divergence can even happen when restricted gene flow is occurring (Allendorf and Phelps, 1981), making the degree of subdivision into locally distinct populations varying. Allele frequencies vary over time and space as populations are of finite sizes (Waples, 1989), but subpopulations should be temporally stable over short time scales when studying population structure.

Freshwater is highly fragmented, consisting of non-continuous water masses that are structured into lakes and rivers of different orders. This leads to fishes, and other freshwater organisms, being structured into different gene pools, often with limited gene flow (Taylor, 1991). Brown trout (*Salmo trutta*) is a species that often form local subpopulations across a diversity of environments (Ferguson, 1989), and are therefore good study subjects concerning local adaptation (Taylor, 1991; Wright, 1931).

Brown trout live in a variety of habitats such as streams, rivers, lakes and costal waters, and is widely distributed. Naturally the species is spread from the northern artic

oceans of Norway to the Atlas Mountains region in the south, from Iceland in the west to western Asia in the east (Behnke, 1986; Elliot, 1994), and it has been successfully introduced by man in at least 24 countries outside its natural range (Elliot, 1994). It is one of the most studied fish species, being of great interest both commercially and recreationally. Due to the considerable variability and plasticity in many aspects of the trouts morphology, ecology and behavior, the one polytypic species goes under many names (about 50 classificated since 1758 (Behnke, 1986)). This variability can be of either genetic or environmental (genetic plasticity) origin; or more commonly a combination of the two (Ferguson, 1989).

In the northern hemisphere spawning activity has been recorded in all months spanning from October to March, with peak seasons for most populations November to December (Elliot, 1994). Spawning takes place over clean gravel in slow running fresh water where the trout carefully select spawning sites. The gravel should be of certain sizes (2-3 cm in diameter), but larger fish can spawn on somewhat larger bedding (Elliot, 1994). After hatching in the spring, the alevins (larvae with yolk sack) remains in the nest until the yolk sack has been used up, then emerge from the gravel as fry. This is a critical stage with high density-dependent mortality when they hastily have to start feeding and establish feeding territories (Elliot, 1994). The trout nearly always spend the first year of its life cycle within the natal stream, often close to the redd location, before they start moving to other, perhaps deeper, parts of the river, to the parent river, to a lake or to an estuary (Elliot, 1994). Brown trout may spawn many times and live for many years (Elliot, 1994).

The trout is found to be genetically differentiated between subpopulations. The genetic isolation observed between populations can come from reproductive isolation in form of a migration barrier (Bouza *et al.*, 1999; Carlsson *et al.*, 1999; Hindar *et al.*, 1991; Ryman, 1983; Taylor *et al.*, 2003). Downstream movements are almost always possible whereas upstream movements may be possible under certain circumstances at some migration barriers (i.e floods) (Carlsson and Nilsson, 2001, Carlsson *et al.,* 1999, Morán *et al.*, 1995). Even if genetic heterogeneity among populations often is associated with restricted gene flow in the form of physical barriers between the demes, differentiation may also occur where no such barriers are present, as between closely situated streams (Carlsson *et al.*, 1999; Morán *et al.*, 1995; Skaala, 1992), within single streams (Carlsson and Nilsson, 2000) and within lakes (Ferguson and Manson, 1981; Ryman *et al.*, 1979). Still, relatively few have studied genetic differentiation within rivers on a small spatial scale (Carlsson and Nilsson, 2000). Restricted movement of resident fish (Bachman, 1984; Torgersen, 2006) due to high degree of local site fidelity, short distances between home areas and spawning grounds (homing) (Carlsson and

Nilsson, 2001; Carlsson and Nilsson 2000), as well as a high tolerance to the presence of related individuals (Griffiths and Armstrong, 2002) may contribute to genetic diversity on a microgeographic scale in populations potentially linked by gene flow.

The geographical distribution of a species are typically more extended than an individuals dispersal capacity, and populations are often genetically differentiated through isolation by distance (i.e. populations in close proximity are genetically more similar than distant populations). Isolation by distance patterns is seldom found in brown trout. Considerable differentiation exists on a micro-geographical scale, and it is shown that individual populations can contain less than one-third of the genetic diversity of the species (Ferguson, 1989). It is argued that the high genetic diversity within the specie is the main reason for the observed population substructure, (Crozier and Ferguson, 1986; Ryman, 1983) as brown trout populations from closely situated streams can be more differentiated than trout separated at greater distance, not supporting the isolation by distance correlation. However, results are conflicting, when isolation by distance also is reported for trout (Carlsson *et al.*, 2000; Carlsson *et al.,* 1999; Estoup *et al.*, 1998b), indicating that spread from single populations over geographical distances can occur.

Genetic variability, both within and between populations, could enhance fitness within a particular habitat, promote colonization and increase distributional range, when more diverse genetics can allow for persistence across a wider range of environments (Carvalho, 1993). For better understanding of how genetic diversity is distributed in fish populations, information not only of the level of genetic variability is needed, but also the relative proportions of variability contained within and between populations (Carvalho, 1993; Ryman, 1983).

This study addresses the genetic structure of naturally produced brown trout on a small spatial scale among and within two river basins in a stable area. The rivers are separated by a small waterfall that may prevent migration, but no migration barriers are known within the rivers. Sixteen microsatellite loci reported to work for brown trout were obtained in order to investigate genetic variation in eight sample sites, one which was represented with temporal replicated samples. The aim was to test if the putative populations were genetically differentiated and on how small scales the trout can be structured into different subpopulations.

Materials and methods

Study area

The two study rivers, Bellbekken and Julussa, are located in the south eastern part of Norway, see Figure 1. Eight different sites have been sampled, three in Bellbekken and five in Julussa; where the distance between the sample sites vary; the shortest distance between two sites is ~50 m (J2 – J3) and the longest is ~7491 m (J1 - B3) (Figure 1). See Table 1 for sample site information.

The lower 1504 m of Bellbekken (including station B1) has been part of a capture-mark-recapture project that started in 2002. Bellbekken is a small forest stream, where reported mean width was 4.46 m and mean depth 21.0 cm during low summer flow conditions (Torgersen, 2006) (based on data from the downstream part). It is under influence of high spring flows and during other seasons also rapidly responding to rainfall as the water levels rise quickly. The stream is naturally stabilized by forest vegetation which provides cover and allochtonus food for fish (Torgersen, 2006). It is subdivided into fast and slower running rapids, deeper sections and pools, and at the outlet to Julussa the stream drains into a small waterfall which is believed to hinder upward migration under most conditions, except possibly when the water level is very high (Figure 1).

Downstream movement is reported as marked trout from Bellbekken have been recaptured in Julussa, but upstream movement is not documented (Olsen and Vøllestad, unpublished data). Brown trout probably colonized Bellbekken before isostatic uplifting of landmasses made the stream inaccessible due to the waterfall. As an alternative, man could have carried them over from nearby streams at a later point of time (Olsen and Vøllestad, 2001a). Brown trout in Bellbekken are small sized (Table 1), rarely reaching 20 cm in length and seldom an age more than 5 years (Torgersen, 2006). It is the only fish species in Bellbekken, except for a few observations of Alpine bullhead (*Cottus poecilopus*).

Julussa is a larger and longer river. It has a broader composition of fish species and holds alpine bullhead at all the sampled sites. Burbot (*Lota lota*), minnow (*Phoxinus phoxinus*), pike (*Esox lucius*) and perch (*Perca fluviatilis*) were also occasionally caught in Julussa. Background information about Julussa is not available, as no previous studies have been conducted in the river.

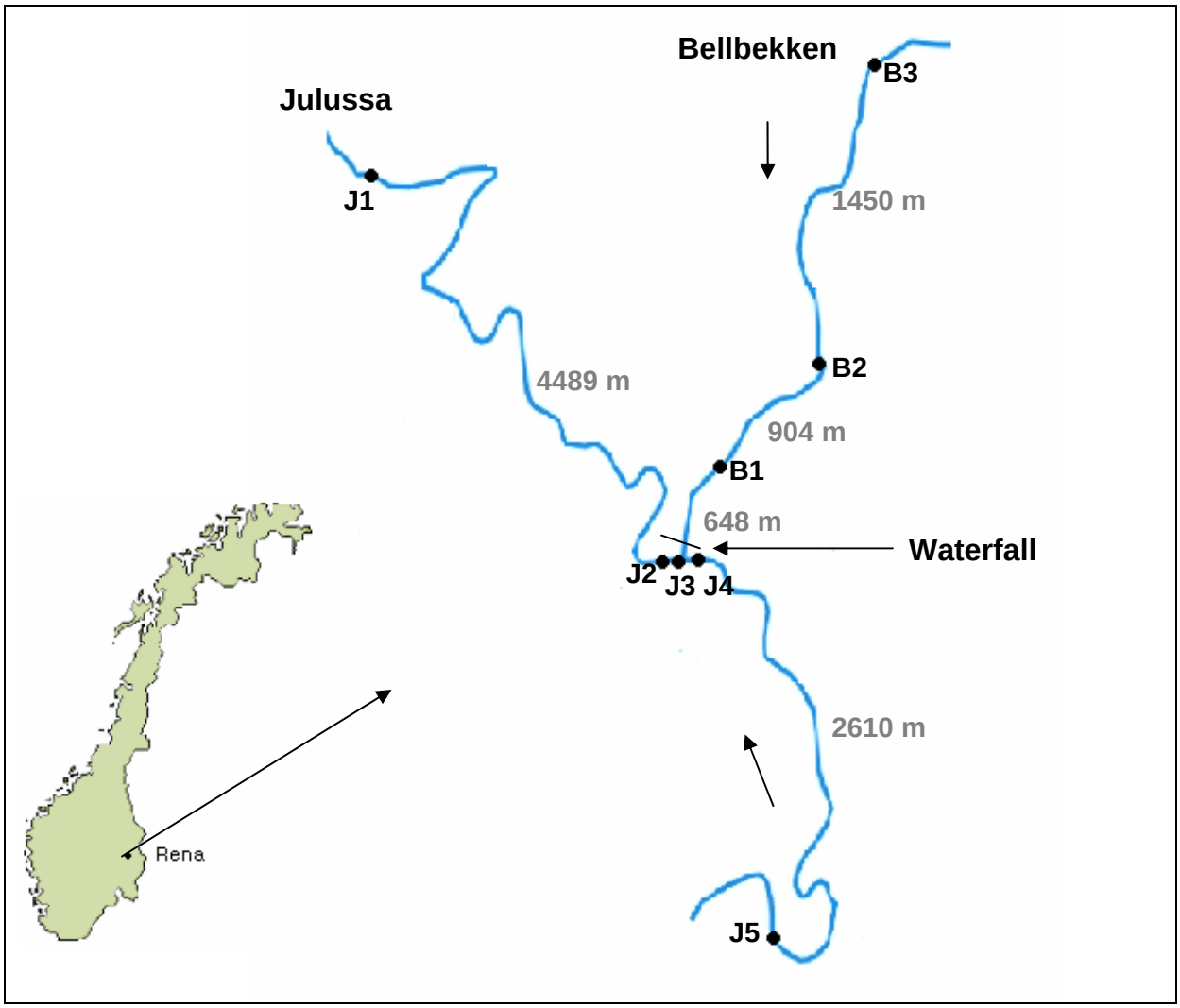


Figure 1. Map of Norway showing the location of the study area, and an enlarged map showing the two sampled rivers; Julussa and Bellbekken, with the sampled sites indicated. Distances (m) between adjacent sample sites and direction of water flow (arrows) are shown.

Table 1. Sample site information. Name of sample site (J; Julussa, B; Bellbekken), sampling period (month and year), sample size (n), length of station (in m) and length of fishes (in cm) (mean ± standard deviation).

Name	Sample period	n	Length station	Length fish
J1	June 2006	61	~200 m	9.10 ± 2.85
J2	September 2005	29	~ 50 m	13.30 ± 3.75
J3	September 2005	36	~ 50 m	9.40 ± 4.71
J4	September 2005	9	~ 50 m	12.80 ± 2.50
J5	September 2006	51	~100 m	11.80 ± 4.98
B1	September 2005	67	300 m	8.20 ± 2.67
	June 2006	70	300 m	9.00 ± 3.51
	September 2006	66	300 m	10.50 ± 3.00
B2	September 2006	50	150 m	11.10 ± 3.66
B3	June 2006	48	~200 m	10.60 ± 4.16
Total:		**487**		

See Figure 1 for the positions of the different sample locations.

Data collection and treatment of samples

Sampling procedure and phenotypic measurements of the fish

Sampling was performed in three different sessions; September 2005, June 2006 and September 2006 (Table 1). The aim of the sampling at each of the different localities was to get approximately 50 individual fish. Difficult sampling conditions in J4 (deep sections, fast running water) resulted in only 9 individuals sampled from this station. J2 and J3 also had small sample sizes (29 and 36 trout respectively), and in addition to the short distances between these three sites (J2, J3 and J4) it will be evaluated if these three stations can be grouped together as one station. B1 was sampled in all three sessions, giving a temporal sample set (Table 1).

Samplings were performed under good sampling condition (i.e., low water flow, stable weather conditions). Each site was systematically electrofished using a Backpack electroshocker (S.Paulsen, Trondheim, Norway), starting at the lower section and sampling upstream (Bohlin *et al.*, 1989) until enough brown trout had been collected (~50). After each round of sampling the fishes were anaesthetized with benzocaine before measuring the fork length (LF, with an accuracy of 1mm). A tissue sample (adipose fin or a small piece of the tailfin) was taken from each fish and stored individually in 96% ethanol. After handling, the fish was put in a bucket of fresh stream water to recover. Handling mortality for the fish was low, as about 1 % of the trout died during electrofishing and handling. All the fish were released within the same station as they were captured.

Genetic analyses

All genetic analyses were performed at the CEES lab at the Department of Biology, University of Oslo.

Isolating DNA

DNA was extracted using the salt-extraction method developed by Aljanabi and Martinez (1997). After the DNA was isolated, the concentration of each sample was checked with NanoDrop (NanoDrop Teqnologies INC) before being diluted with dH_2O to a concentration of <20 ng/mL.

PCR

Sixteen microsatellite DNA markers were analysed; seven dinucleotides and nine tetranucleotide, (Appendix 1). Microsatellites studied were CA040261, CA053037 (Vasemagi *et al.*, 2005), MST-85 (Presa and Guyomard, 1996), SSA-85 (Oreilly *et al.*, 1996), STR2INRA (Estoup *et al.*, 1998a), Strutta-12 (Poteaux *et al.*, 1999), TAP2B (Grimholt *et al.*, 2002) (dinucleotide), CA060177 (Vasemagi *et al.*, 2005), SSaD58, SSaD71, SSaD157, SSaD190, SSaD237 (King *et al.*, 2005), SsaD85 (Eackles and King, Unpublished), SSaD170 (King *et al.*, 2005) and SSsp2213 (Paterson *et al.*, 2004) (tetranucleotide). Three of the primers, SSaD71, SSaD85 and SsaD170, could be run together on the same PCR programme as a triplex, and CA060177 and TAP2B could be run together as a duplex.

The amplification reaction was performed with a total volume of 10 μL for the eleven simplexes and the duplex, and 15 μL for the triplex. Each sample contained a variation of the reagents for a standard polymerase chain reaction; 1-1.5 μL DNA, 0.200-0.500 μM of each primer (Applied Biosystems, DNA technology A/S), 0.300-0.450 Mm dNTP, 10X 1-1.5 PCR buffer (Promega); $MgCl_2$ optimized for each locus, and 0.030-0.050 units of Biotaq DNA polymerase (Bioline). (See Appendix 2 for individual setup). For each multiwell plate, a dH_2O sample was included as a negative control.

With two (CA040261 and CA053307) and three (SSaD58, SSaD237,Triplex) of the primer setups being able to use the same PCR programs, a total of ten different programs for amplification were run, were the duplex followed a touchdown program. The triplex and the simplexes were run on a PCR machine with the following profile: denaturating temperature of 94-95˚ C for 2-5 minutes, followed by 30-35 cycles of denaturation of 92-95˚C for 30-45s, a primer annealing step of 55-66˚C for 30-45s and an extension step of 68-72˚C for 30s-2 minutes. Finally, an extension step at 68-72˚C for 5-10 minutes were used. The touchdown program used for the duplex consisted of a denaturation step at 94˚C for 2 min. Then 20 cycles with a denaturation of 94˚C for 30s, the annealing temperature started at 60˚C and was increased by 0.5˚C for each cycle, extension step for 30s at 72˚C. The rest of the 15 cycles comprised of denaturation at 94˚C for 30s, annealing temperature at 60˚C for 30s and extension at 72˚C for 30s. Finally an extension step at 72˚C for 5 minutes. (See appendix 3 for individual setup of the different PCR programs).

Five different PCR machines were used for PCR amplification: Biometra T‧ gradient thermocycler (Biometra), Biometra T‧ Thermoblock (Biometra), Mastercycler Ep (Eppendorf), PTC-0200 DNA engine (MJ research, Waterton, MA, USA), and a PTC-240

DNA engine Tetrad 2 cycler (MJ research, Waterton, MA, USA). The duplex was only run on the Mastercycler Ep.

Genotyping

Two different genotyping sets were made from the 16 Microsatellite PCR products due to their colourings and length (see Appendix 4). The first set contained 2 µL of a mixture consisting of 2 µL of SSa-85, SSsp2213, 3 µL of CA040261 and the triplex (SSaD71, SsaD85, SsaD170), 4 µL of the duplex (CA060177, TAP2B) and CA053307. The second set contained 2 µL of a mixture out of 2 µL SSaD190, 3µL of SSaD58, SSaD157, SSaD237, MST-85, 4 µL of STR2INRAand 5 µL of Strutta-12. PCR products were applied to a mixture of 10 µL formamide (denaturation) and 0.125 µL allelic standard ladder (Gene scan- 500 LIZ, Applied Biosystems, CA, USA), which subsequently was denatured for 3 minutes at 95˚ C (parts the DNA treads), directly stored on ice for 3 minutes (leaving the treads unpaired) before being run on an 3730 DNA Analyzer (Applied Biosystems) according to the manufactures recommendations. The software Genemapper (Applied Biosystems) was used to inspect and revise the microsatellites manually. Twenty samples were genotyped twice to estimate scoring errors.

A total of 487 individuals were analysed successfully. From this data two separate datasets were made; one spatial dataset including all the stations in Julussa and Bellbekken (J1, J2, J3, J4 and J5, B1S06, B2 and B3), and one temporal dataset including three samples from B1 (B1F05, B1S06 and B1F06) (Table 1). B1 has been part of an earlier capture –mark –recapture project (from 2002), therefore this sample site had a large amount of available fish to include in this study. No fish were included twice in the dataset due to individual tagging of the fish and visual identification of young of the year due to length. This corresponds to a sampling method where the fish would be taken out of the system (destructive sampling). Total sample size for the spatial dataset were 354 and for the temporal dataset 203 individuals. Except where specifically noted, all results given are from analysis of the spatial dataset.

Analyses

Within population patterns of genetic diversity

The number of reproducing individuals combined with the number of immigrants received from other populations determines the amount of genetic diversity found in the population. Genetic diversity indices were calculated by FSTAT Version 2.9.3 (Goudet, 1995; Goudet, 2001) and GENALEX, Version 6 (Peakall and Smouse, 2006). FSTAT was used both to calculate the number of alleles per locus and to calculate number of alleles per sample. The number of private alleles was calculated using GENALEX. Private alleles are alleles unique to a given sample. Tests for significant deviations from Hardy-Weinberg equilibrium were performed for each locus and population combination using an exact test were the p-values were estimated without bias, using a Markov chain method following the algorithm of Guo and Thompson (1992). This test is implemented in GENEPOP version 3.1.c. (Raymond and Rousset, 1995) (Markov chain parameters set to default values). (Corrections for multiple tests were performed by applying a Bonferroni correction (Rice, 1989)).

Observed and expected heterozygosity of the markers were also estimated using GENEPOP. Linkage disequilibrium for all pairs of loci within each population was tested with FSTAT. This is a term used for a non random association (more or less frequent than expected) of two or more alleles, resulting in less information from the loci involved. (Bonferroni corrections were applied (Rice, 1989)).

In order to test for the presence of null alleles or stuttering when scoring the alleles, MICRO-CHECKER, Version 2.2.3., was used (Van Oosterhout *et al.*, 2004). The application uses a Monte Carlo simulation (bootstrap) method to generate homozygote and heterozygote allele size difference frequencies, before the Hardy-Weinberg theory of equilibrium is used to calculate expected allele frequencies and the frequencies of any null allele detected. Null alleles are the consequence of one or more alleles failing to amplify during PCR, thereby creating a false number of homozygote in the data (homozygote excess). This will then bias the estimate of allele frequencies. Stuttering is caused by slight changes in the allele size during PCR, due to a haltered polymerase activity. This gives small peaks of allele fragments that are formed in different lengths from the true allele by an integral number of repeats. This is a phenomenon more common in dinucleotides (Edwards *et al.*, 1991) compared to loci with larger repeated motifs.

Among-population patterns of genetic diversity

The degree of genetic uniqueness are determined by the combined effect of the number reproducing individuals within populations, the number of migrants exchanged among populations, and the differences in environmental conditions experienced by the different populations. Estimates of genetic differentiation between sample sites were analysed by pairwise F_{ST} values, using Weir and Cockerham's (1984) estimator θ, as calculated in FSTAT version 2.9.3. (Goudet, 1995; Goudet, 2001). F_{ST} is the fraction of total genetic diversity at the molecular level attributable to between-population differences. It was tested, due to the short distance between sites and low sample sizes, if J2, J3 and J4 could be grouped together. A test for overall population differentiation (not assuming Hardy-Weinberg equilibrium) was performed. The null hypothesise of no spatial structure (i.e; average $F_{ST} = 0$) was tested by a permutation test. In this test all individuals are reassigned to a sample (permutation and resampling of multilocus genotypes among pairs of samples) at random before θ is recalculated for the rearranged dataset (Weir and Cockerham, 1984). Permutations were repeated 3000 times, and the probability of the null hypothesis was taken as the proportions of replicates that yielded an estimate (θ) of F_{ST} as high, or higher, than the observed levels of population differentiation (Goudet *et al.*, 1996). The degree of population subdivision from multilocus estimates of F_{ST} (Weir and Cockerham, 1984) for all population pairs (Goudet, 1995) was determined.

The F_{ST} values for the temporal dataset were estimated together with the F_{ST} values for all sample sites in order to confirm temporal stability. It is important to substantiate temporal stability in order to conclude that the observed F_{ST} values reflect an actual spatial differentiation and not temporal heterogeneity or any sampling artefacts. The estimation was done following the randomizations described above (28 000 permutations). Bonferroni corrections were applied (Rice, 1989).

Isolation by distance is the correlation between the pairwise multilocus F_{ST} estimates and the waterway distance between the different sample locations. The idea is that individuals may be spatially distributed across some region with local dispersal; the result in this situation is that the allele frequencies will vary gradually across the region; with the populations further apart being more isolated from each other than the populations closer together.

The geographical distance between the sites was estimated using the Geographical Information System (GIS), following the river contours. A Mantel test implanted in GENEPOP, version 3.1.c. (Raymond and Rousset, 1995) was used. The null hypothesis, that the regression slope is zero, is tested by computing a regression of $F_{ST} / (1 - F_{ST})$ (Rousset,

1997) against the geographical distance before performing a Mantel test (Mantel, 1967). This test was designed to test for the independence of the elements of two matrices (pairwise genetic and geographical distances between pairs of samples in this case). The method calculates a statistics closely related to the correlation coefficient for each pair of matrices by a Monte Carlo resampling method. The value calculated on the real dataset is then compared, and by holding one matrix constant and vary the other, both in rows and columns, the significance of the correlation on the initial matrices can be tested. The number of permutations was set to 1000.

Assignment of individuals and estimates of population structure

Genetically structured populations can be viewed as a set of discrete subgroups, each where the alleles has distinct frequencies. Migration can lead to gene flow, thereby opposing genetic differentiation since this constantly bring new gene material into the population. Only a few individuals migrating to a population with reproductive success each generation will be enough to prevent genetic subgroups to form (Spieth, 1974).

Two different analyses were used in order to test if the samples collected in this study could be viewed as genetic structured, or if gene flow between them is high enough to hinder substructuring. Geneclass2 (Piry *et al.*, 2004) was used for an individual assignment test and STRUCTURE, Version 2.2. (Pritchard *et al.*, 2000; Pritchard *et al.*, 2007) was used to get an approximation of *K*, the number of separated populations. In GeneClass2, un-sampled populations can be taken into account, whereas STRUCTURE assumes that the true candidate population is included in the analysis.

To determine the probability of individuals being assigned to populations other than the population of origin, an individual assignment test (self assignment, (Hansen *et al.*, 2001)) was performed (Paetkau *et al.*, 1995). This test, first developed by Paetkau *et al*. (1995), calculates the probability of each individual belonging to a certain sample, where the computation is based on the individual's multilocus genotype compared to the allele frequencies of the specific sample. The individuals are then assigned to the sample where they have the highest probability of belonging. The software GeneClass was used for these calculations, and as recommended by Cournet (1999), a partly Bayesian approach was used, specifically the method of Rannala and Mountain (1997) (see also Hansen et al. 2001 for review).

One of the challenges with assignment test is distinguishing between residents (individuals born in the population where they were sampled) that are misassigned because

they have a genotype that is most likely to occur in another population by chance, and first generation immigrants who are misassigned because they were born somewhere else than where they were sampled (Paetkau *et al.*, 2004). To overcome this problem one can use a Monte Carlo resampling method and identify statistical thresholds beyond which individuals are likely to be first generation immigrants (Cornuet *et al.*, 1999; Paetkau *et al.*, 2004). The principle is to approximate the distribution of genotype likelihoods that would be found in a particular group of "real" resident individuals, and then compare the likelihood calculated for sampled individuals to that distribution. The Monte Carlo resampling method of Paetkau et al. (2004) was used. In this method a simulated individual is obtained by drawing, with replacement, multilocus gametes from randomly chosen individuals in each population of the dataset. To decrease bias, the "leave-one-out" option was included. This means that the individual being assigned is not included in the estimation of allele frequencies; its genotype is subtracted from the allele distribution in which they were included when performing their assignment (Paetkau *et al.*, 1998). The number of simulated individuals was set to 1000 (which means that if the genotype of an individual occurred once in a population in the 1000 simulations, the estimated probability of origin from this population would be 1/1000).

Different assignment thresholds of 80 and 90% were used in order to assign individual to the different stations (Berry *et al.*, 2004; Paetkau *et al.*, 2004). The assignment was done according to the exclusion method (Cornuet *et al.*, 1999) where the individual is assigned to a population according to where it has the highest probability of belonging. In this method it is not required that the true population of origin has been sampled, as it does not compare populations but treats each one separately. This may be an advantage when it is not possible to sample all the candidate populations (Cornuet *et al.*, 1999). In addition, Cornet *et al*.(1999) warn about two overlapping types of errors concerning exclusion tests with Monte Carlo resampling; type A errors when the correct population is not listed; resulting in individuals being assigned to two or more populations with different probabilities, and type E errors when the correct population is listed with additional possible populations; resulting in confusing assignments when individuals are assigned to more than one population with the same probability. Trying to minimize these errors, individuals were assigned to "river" (Julussa or Bellbekken) when two or more sample sites within one river were being a potential "home" for a specific individual (two or more sample sites in either Julussa or Bellbekken having higher probability than 80 or 90% for one individual). When stations from both rivers could be regarded as potential "home" to an individual, the individual was classified as "unassigned", as were individuals not reaching the assignment threshold for any populations

or rivers. Misassignments with high probabilities (<80%) are unlikely to occur from a random combination of alleles (Zamudio and Wieczorek, 2007), and individuals assigned to a population other then to the original sampled were deduced as migration events (or offspring of recent immigrants).

To test for the number of separate genetic units constructed by the analysed samples, a model-based Bayesian procedure was used with the programme STRUCTURE. This method assumes Hardy-Weinberg- and linkage equilibrium within populations and attempts to find groups of population that are not in disequilibrium. The method does this by identifying clusters of individuals based on their genotype at multiple loci without prior knowledge about their population affinities. The model assumes *K* genetic clusters (set manually), each characterized by a set of allele frequencies at each locus; the admixture model then probabilistically estimates the proportion of individuals with ancestry in each cluster and estimate the log of probability of data Pr(X|K) for each value of *K* that is given. A quantification of how likely each individual is belonging to each group under consideration is also given (Q), information that can then be used to assign individuals to populations. The membership coefficients (Q) sum to one for each individual, as STRUCTURE assumes that the true population of origin has been sampled

A series of pilot runs were used in order to estimate Pr(X|K) were X represents the data, for *K* between 1 (the expected value if all populations belonged to the same population) and 8 (the expected value if all samples collected belonged to one population) to get an idea of *K*. The model choice criterion called 'Ln P(D)' in the STRUCTURE output is obtained, and the maximum value (were LnP(D) stops increasing or increase slower) is believed to identify the *K* that best describes the data. Most of the parameters were set to default values as advised in the user's manual of STRUCTURE (Pritchard *et al.,* 2007). Specifically, the admixture model and the option of correlated allele frequencies were used, as this is considered the best option by Falush et al (2003) in the case of subtle population structure. As the model could produce different likelihood values for each value of *K*, 20 independent analyses were run for each *K*, using 1 500 000 interations (following a burn in-period of 700 000) (Pritchard *et al.*, 2000). Output files obtained from STRUCTURE were summarized in R (http://www.r-project.org/index.html) as described by Ehrich (2006).

From the initial runs it was determined that the true value of *K* (with the highest values of Ln P(D)) fell between 2 and 4. Focusing on this lower range of *K* for more detailed analyses, 25 replicates were run using the same parameters as the pilot study. In the user's manual the authors (Pritchard *et al.*, 2007) warn about the difficulties of estimating the true *K*

due to computational difficulties and the complications of the biological interpretation of estimating the number of populations. This can lead to an overestimation of assumed populations, and therefore ΔK were calculated (Evanno *et al.*, 2005). ΔK is an ad hoc quantity based on the rate of change in the log probability of data between successive *K* values (it takes account for the shape of the log likelihood curve with increasing *K* and variance among estimates in multiple runs). Being an estimate of the increase between *K* and *K*-1 this value is not computed for *K*=1 (Evanno *et al.*, 2005). ΔK was then compared together with Ln P(D) to estimate the "true" number of *K*. In order to strengthen the liability that STRUCTURE (Pritchard *et al.*, 2000) had "correctly" calculated the number of clusters in the two rivers, analyses of single rivers (Bellbekken *K* set from 1 to 3, Julussa *K* set from 1 to 3) were performed using the same parameters as the previous run to see if the resulting *K* 's added up to the resulted *K* from the analysis of both rivers combined. A series of runs containing prior sampling location information (USERPOPINFO = 1) were also computed for the whole dataset, and for the different rivers, in order to check for different results.

Individual and population membership coefficients of ancestry (Q) in the inferred populations were graphed in the program DISTRUCT version 1.1 (Rosenberg, 2004) to easier visualize the outcome from STRUCTURE outputs (with USERPOPINFO=1, as required in DISTRUCT). Here the estimated quantifications (Q) for belonging to different subpopulations (*K*) are represented as colours, and individuals are depicted as bars that are partitioned into coloured segments according to the membership coefficients in the subgroups.

Results

Within-population patterns of genetic diversity

No scoring errors were recorded for the 20 samples that were scored twice. Missing data per locus per population ranged from 6% (ssaD237, B2) to 100% (Strutta-12, J3). Considerable variations in the microsatellites studied were observed when the number of alleles detected at each locus varied from three (TAP2) to 21 (SSaD157 and SSaD337) (Table 2). For each population the number of alleles per sample ranged from two alleles in TAP2 (J4) to 18 alleles in SSaD157 and SSaD237 (J1, B1S06, B1F06). In total 166 alleles could be detected from the 16 loci (mean of 10,375 alleles per locus). Eight Private alleles were found in 13 individuals (Table 2), for all the samples together.

Thirty of 160 (18.5%) tests for deviation from Hardy-Weinberg equilibrium were significant before Bonferroni corrections (Rice, 1989). After the corrections, eleven tests still showed significant signs of being in Hardy-Weinberg disequilibrium, where the CA050307 locus counted for nine of them (Table 2, marked in bold). The presence of null alleles at loci CA050307 is the most likely explanation for this deviation, as MICRO-CHECKER (Van Oosterhout *et al.*, 2004) also suggest (see below). Expected gene diversity (H_E) averaging across 14 loci (CA053307 and Strutta-12 were left out due to null-alleles and low scoring results) varied from 0.694 (J4) to 0.767 (J1), whereas observed heterozygosity varied from 0.687 (B3) to 0.768 (B1F06) (Table 2). The linkage disequilibrium test confirmed that the loci were independent. After 1260 interactions, only 72 (5.71%) loci comparisons were significant at the 0.05 level (temporal dataset also included). None were significant after Bonferroni corrections (Rice, 1989).

After analysing the data from MICRO-CHECKER following Østbye *et al.* (2006) it was found that the differences between the observed allele frequency and the adjusted allele frequency were minor in most cases. In total, 19 null alleles were suggested in of 96 tests of locus-x-population combinations. J1 had three locus where null alleles were suggested (CA040206, CA053307, STR2), J2 one (CA053307), J5 two (CA053307, STR2), B1 one (CA053307), B2 one (CA053307) and B3 three (CA040206, CA053307, SSaD170). The locus-specific presence of null alleles was; CA040261 (3 cases), CA053307 (12), SSaD170 (1) and STR2 (3). Of these 19 null alleles, CA053307 had the majority with 12 cases. CA053307 was excluded from further analysis, both due to null alleles, but also because of suspicion of stuttering that might have resulted in scoring errors. Stuttering is reported for

CA053307 in MICRO-CHECKER. The other reported loci with null alleles were considered unlikely to alter the results, and was kept for the following analysis. Due to low scoring results in Strutta-12 for population J3 and J5 (100% and 80% respectively); this locus was excluded from the analyses.

Table 2. Summary of microsatellite data. Sample site (localization) and sample size is shown together with a summary table of each population and loci; observed number of alleles per locus (Total no. alleles) and per population (Na.), (private alleles in parenthesis), expected (H_E) and observed heterozygosity (H_O) and number individuals with successfully amplified locus (n). Mean H_E and H_O for all loci in each population is indicated at the bottom (CA053307 and Strutta-12 left out in the estimate). Sampling site information is shown in Table 1.

Localization		***J1***	***J2***	***J3***	***J4***	***J2-J4***	***J5***	***B1F05***	***B1S05***	***B1F06***	***B2***	***B3***
Sample size		61	29	36	9	74	51	67	70	67	50	48
Locus	**Statistics**											
CA040261												
Total no.	Na	8	8	7	5	8	7	9	9	10	9	9
alleles: 10	H_E	0.779	0.715	0.749	0.672	0.740	0.765	0.849	0.829	0.824	0.817	0.799
	H_0	0.600	0.607	0.629	0.875	0.648	0.840	0.812	0.740	0.667	0.831	0.923
	n	55	28	35	8	71	50	65	69	66	50	42
CA053037												
Total no.	Na	6	5	5	3	5	6	7	6	6	5	5
alleles: 6	H_E	**0.627*** **	**0.530** **	0.584	0.653	**0.589*** **	**0.571*** **	**0.724*** **	**0.540*** **	**0.702*** **	**0.739*** **	**0.736*** **
	H_0	0.418	0.273	0.545	0.667	0.459	0.341	0.361	0.255	0.400	0.468	0.500
	n	55	22	33	6	69	44	48	61	53	47	45
CA060177												
Total no.	Na	7 (1)	6	5	3	7	6	6	7	7	5	7
alleles: 8	H_E	0.712	0.724	0.694	0.648	0.707	0.705	0.692	0.694	0.673	0.734	0.697
	H_0	0.717	0.724	0.781	0.500	0.725	0.640	0.642	0.714	0.721	0.746	0.690
	n	46	29	32	8	69	50	63	67	59	49	43
MST-85												
Total no.	Na	7	7	7	5	7	7	7	6	7	6	6
alleles: 7	H_E	0.786	0.803	0.765	0.750	0.785	0.757	0.724	0.710	0.565	0.710	0.754
	H_0	0.817	0.846	0.767	0.875	0.813	0.744	0.743	0.620	0.694	0.627	0.714
	n	60	26	30	8	64	43	67	70	62	50	46
SSa85												
Total no.	Na	4	4	4	4	4	4	4	4	4	4	4
alleles: 6	H_E	0.740	0.744	0.740	0.698	0.746	0.741	0.685	0.672	0.582	0.695	0.675
	H_0	0.721	0.643	0.706	1,000	0.718	0.860	0.696	0.740	0.521	0.697	0.625
	n	61	28	34	9	74	50	66	69	65	50	48
SSaD58												
Total no.	Na	6	5	7	5	6	11 (1)	5	7	5	5	6
alleles: 11	H_E	0.687	0.633	0.657	0.630	0.650	0.712	0.614	0.571	0.648	0.617	**0.621***
	H_0	0.714	0.517	0.694	0.778	0.635	0.633	0.647	0.460	0.545	0.582	0.730
	n	56	29	36	9	73	49	67	68	64	50	44
SSaD71												
Total no.	Na	7	5	6	3	6	5	7	7	7	6	6
alleles 8	H_E	0.718	0.668	0.665	0.611	0.668	0.672	0.755	0.747	0.701	0.767	0.734
	H_0	0.750	0.793	0.657	0.778	0.726	0.680	0.757	0.620	0.617	0.761	0.773
	n	60	29	35	9	73	50	67	70	67	50	47

SSaD85												
Total no.	Na	12	10	11	7	11	11	11	10	11	10	12
alleles: 12	H_E	0.853	0.797	0.863	0.769	0.84	0.837	0.86	0.865	0.856	0.862	0.857
	H_0	0.867	0.759	0.886	0.778	0.822	0.860	0.900	0.900	0.851	0.896	0.848
	n	60	29	35	9	73	50	67	70	67	50	47
SSaD157												
Total no.	Na	18 (2)	12	13	7	14	14 (1)	11	12	14	10	12 (1)
alleles: 21	H_E	0.884	0.879	0.878	0.827	0.887	0.891	0.843	0.819	0.705	0.847	0.844
	H_0	0.881	0.923	0.882	0.889	0.899	0.851	0.871	0.837	0.729	0.896	0.922
	n	59	26	34	9	69	47	67	70	65	49	48
SSaD170												
Total no.	Na	11	10	11	6	13	13	11	12 (1)	12	9	13
alleles: 14	H_E	0.862	**0.850*** **	0.87	0.734	0.862	0.88	0.847	0.831	0.871	0.833	0.84
	H_0	0.880	0.760	1,000	0.875	0.900	0.875	0.812	0.760	0.761	0.851	0.788
	n	58	28	34	8	70	48	67	69	67	50	46
SSaD190												
Total no.	Na	4	5	4	4	5	4	4	4	4	4	4
alleles: 5	H_E	0.534	0.609	0.536	0.543	0.575	0.575	0.723	0.693	0.692	0.697	0.710
	H_0	0.557	0.552	0.556	0.556	0.554	0.469	0.662	0.720	0.660	0.646	0.561
	n	61	29	36	9	74	49	65	68	67	50	47
SSaD237												
Total no.	Na	18	17	15	9	18	16	15	18	18	13	14
alleles: 21	H_E	0.893	0.922	0.911	0.858	0.919	0.917	0.894	0.871	0.907	0.876	0.908
	H_0	0.915	0.769	0.861	1,000	0.845	0.840	0.839	0.851	0.929	0.906	0.850
	n	59	26	36	9	71	50	64	62	61	47	42
SSsp2213												
Total no.	Na	7	5	6	5	8	7	5	8	6	7	7
alleles: 9	H_E	0.709	0.757	0.76	0.636	0.768	0.760	0.733	0.751	0.754	0.753	0.712
	H_0	0.759	0.679	0.829	0.444	0.722	0.784	0.667	0.913	0.783	0.682	0.692
	n	58	28	35	9	72	51	66	69	66	46	46
STR-2												
Total no.	Na	10	10	10	9	11	10	10	9	10	8	9
alleles: 12	H_E	0.847	0.845	0.821	0.847	0.844	0.812	0.786	0.756	0.779	0.825	0.793
	H_0	0.638	0.893	0.686	0.857	0.786	0.634	0.758	0.735	0.767	0.852	0.882
	n	58	28	35	7	70	41	54	62	52	49	43
STRUTTA-12												
Total no.	Na	12	10	Na	2	10	7	10	13 (1)	13	10	12
alleles: 5	H_E	0.770	0.770	-	0.500	0.770	0.815	0.755	0.678	0.777	0.775	0.789
	H_0	0.763	0.64	-	1,000	0.654	0.800	0.712	0.714	0.744	0.813	0.841
	n	59	25	0	1	26	10	64	66	64	49	43
TAP2												
Total no.	Na	3	3	3	2	3	3	3	3	3	3	3
alleles: 3	H_E	0.587	0.561	0.552	0.492	0.550	0.552	0.467	0.499	0.314	0.491	0.518
	H_0	0.533	0.724	0.571	0.375	0.611	0.471	0.362	0.510	0.229	0.507	0.619
	n	60	29	35	8	72	51	67	69	64	49	48
All loci												
Total no.	H_E	0.767	0.751	0.757	0.694	0.753	0.765	0.758	0.752	0.757	0.758	0.726
alleles: 166	H_0	0.749	0.738	0.750	0.766	0.743	0.737	0.736	0.759	0.768	0.736	0.687

Deviation from Hardy-Weinberg equilibrium is indicated in bold, ***P< 0.001, ** P< 0.01, *P< 0.05
Tablewide significance levels were applied, using the sequential Bonferroni technique (Rice, 1989).

Among-population patterns of genetic diversity

It was tested, using estimates of pairwise F_{ST} values, if the samples from J2, J3 and J4 could be grouped together. The F_{ST} values were consistently low and nonsignificant (Table 3). The three stations were therefore grouped together for the following analysis as station J2_4.

For the temporal dataset, the pairwise F_{ST} values were low and not significantly different from zero (Table 4, marked in bold). Lack of significant differences between years suggests temporal stability, which means that the observed F_{ST} values between sites is not confounded by temporal heterogeneity, but reflects the actual spatial differentiation. The samples from the station with temporal stability (B1F05, B1S06 and B1F06) could have been pooled together and been analysed as one sample (B1), but since this would more than double the sample size at this station (B1) compared to the others, one sample (B1S06) was chosen to be included in the spatial dataset.

The now six stations and 14 microsatellites revealed significant genetic structure in 23 out of 28 pairwise multilocus F_{ST} estimates after Bonferroni corrections (Table 4). The significant values ranged from 1.0% differentiation between J1 and J5, and up to 5.22% differentiation between J2_4 and B3, with an average measure of genetic differentiation, F_{ST} of 0.023 (without B1F05 and B1F06). The F_{ST} value between J1 and J5 (0.0100) was the only significant differentiation internally in Julussa, whereas in Bellbekken all the differentiations were significant. It was especially interesting to notice that internally in Bellbekken, the smaller stream with shorter distances between sampling sites, the F_{ST} values were larger than those internally in Julussa with one order of magnitude, indicating that Bellbekken is more differentiated than Julussa.

The isolation by distance was tested for by comparing pairwise geographic and genetic distances. No clear geographical pattern was found to the genetic differentiated samples, although a weak positive trend (slope 0.000002) related to a marginally significant p–value of 0.0790 was indicated (Figure 2). From the figure it is easy to see that there is a large variability in the data. The J2_4 and Bellbekken 1, 2 and 3 comparisons (the three diamond points in the upper left in Figure 2) demonstrate short distances with high F_{ST} values, whereas internally in Julussa (indicated as green triangles in the bottom of Figure 2), long geographic distances between sites with correspondingly low F_{ST} values are illustrated.

Table 3. Estimates of genetic differentiation (F_{ST}) between J2, J3 and J4 and significance of F_{ST} values tested. Sampling site information is shown in Table 1.

	Fst	P-value
J2 vs J3	-0.0014	0.249
J2 vs J4	0.0004	0.356
J3 vs J4	0.0004	0.118

Table 4. Estimates of genetic differentiation (F_{ST}) between pairs of samples and significance of F_{ST} values tested for all the six sample sites after J2, J3 and J4 have been grouped into one station; J2_4. Estimates of the temporal dataset are indicated in bold. Sampling site information is shown in Table 1.

	J1	J2_4	J5	B1F05	B1S06	B1F06	B2
J2_4	0.0064						
J5	0.0100***	0.0013					
B1F05	0.0221***	0.0212***	0.0178***				
B1S06	0.0269***	0.0262***	0.0216***	**-0.0002**			
B1F06	0.0220***	0.0212***	0.0156***	**0.0021**	**0.0011**		
B2	0.0318***	0.0331***	0.0279***	0.0067***	0.0051*	0.0038*	
B3	0.0426***	0.0522***	0.0443***	0.0182***	0.0147***	0.0143***	0.0213***

***P<0,001, * P< 0,05
Tablewide significance levels were applied, using the sequential Bonferroni technique (Rice, 1989)

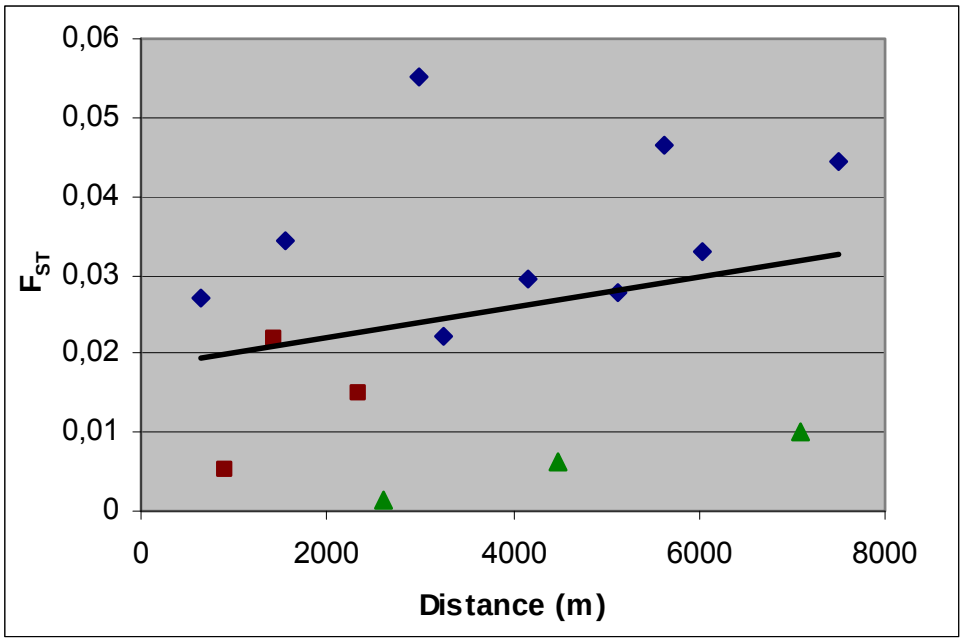


Figure 2. Relationship between genetic differentiation and geographical distance between sample sites. Diamonds (♦) indicate Julussa and Bellbekken comparisons, triangles(▲) indicate within Julussa comparisons, squares (■) indicate comparisons within Bellbekken. The fitted line (b = 0.00002) explains 6.9% of the variation in the data.

Assignment of individuals and estimates of population structure

The assignment tests were run for the six sample sites. The percentage of trout assigned to specific station or to river was calculated under two levels of stringency; 80 and 90% assignment thresholds, and the numbers of separate populations (*K*) in the dataset were estimated. In order to get as solid data as possible, fish with less than 10 successfully amplified loci were removed from the data set (three fishes removed, n = 351 for these analyses) as this could affect the successfulness of the assignment (Berry *et al.*, 2004).

The assignment test (Rannala and Mountain, 1997) with 80% threshold placed from 8 (B2) to 53% (B3) (mean 26%) of the individuals back to the sample locations they had been obtained when the number of simulated individuals were set to 1000 (Figure 3). For comparison, 5 (J2_4) to 45% (B3) (mean 22%) was assigned back to their original sample with an assignment threshold of 90% (Figure 3). Unassigned individuals in the 80% threshold analysis were from 27 (B3) to 50% (B1) (Mean 41%), whereas 52 (B3) to 73% (B1) (mean 63%) of the individuals was unassigned with the 90% assignment threshold. The unassigned individuals did not make the threshold for any population (or river) and were the majority in all, except two cases concerning the 80% assignment thresholds; J2_4 had a higher assignment to "river" (J) and B3 had a higher assignment to self. Individuals assigned to the same river as they were sampled composed from 19 to 41% (mean 23%) for the 80% threshold and from 2 to 23% (mean 10%) with 90% assignment threshold. Julussa, as a total had four individuals assigned to Bellbekken, and Bellbekken had a total of two individuals being assigned to Julussa (both cases at the 80% assignment threshold). (Comparable analyses with 10 000 simulated individuals is given in Appendix 5; no clear differences where detected).

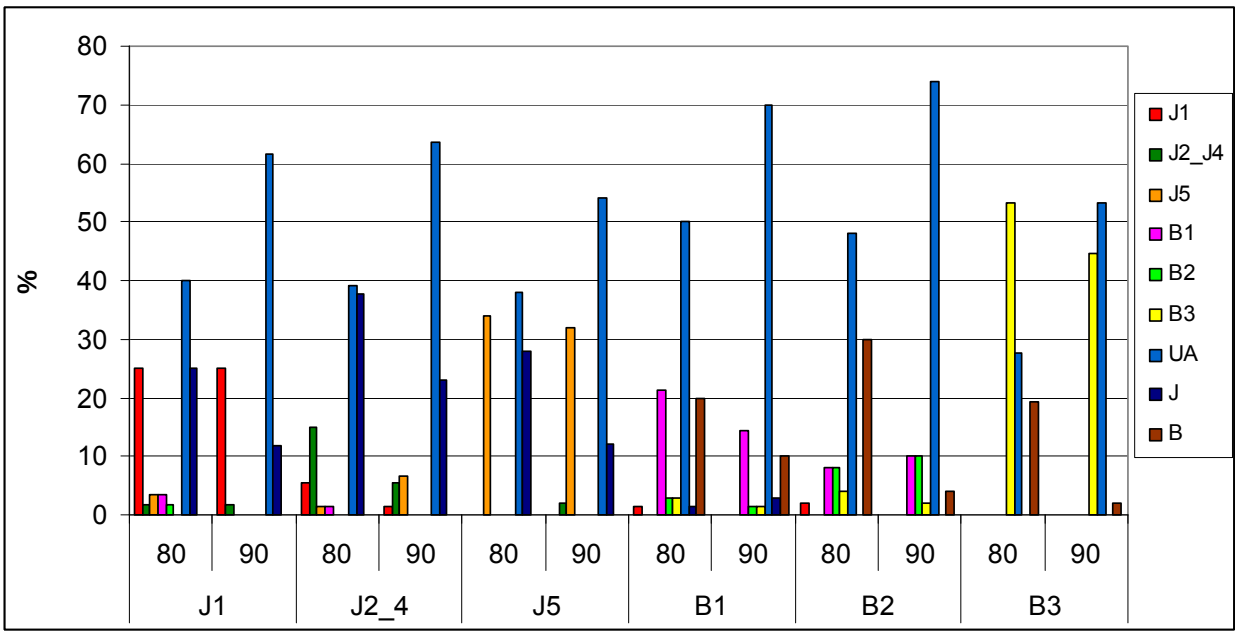


Figure 3. Assignment test with 1000 simulated individuals. The results are given for tests with an assignment threshold with 80% (80) and 90% (90), values reported in percent. Unassigned individuals and individuals assigned Julussa (J) and Bellbekken (B), when individuals were assigned to multiple stations within the same rivers, are reported for each sample site and threshold. Sampling site information is shown in Table 1.

The Bayesian estimate of population structure (Pritchard *et al.*, 2000) showed that $K = 2$ gave a better fit than alternative models when testing on all sample sites (Figure 4). This was evident from both the LnP(D) values (Figure 4a) and ΔK (Figure 4b). Even if the summarization of the LnP(D) show that $K = 2$ has lower values than $K = 4$ (average LnP(D) for $K = 2$: -16227,8, average LnP(D) for $K = 4$: -16205,7), $K = 2$ shows a higher consistency in the LnP(D) values compared to the higher variance in the LnP(D) values computed for $K = 4$ (Figure 4a). This indicate that $K = 2$ is a better estimate for the data (Evanno *et al.*, 2005; Pritchard *et al.*, 2007). ΔK takes this variance into account when this parameter is sensitive for variance in multiple runs and estimates $K = 2$ to be the best description of the data (Figure 4b). That $K = 2$ for both rivers was further supported from the resulted runs from single rivers. The best overall result for both rivers run separately was $K = 1$ (Figure 4c, d).

Running STRUCTURE with additional information concerning sample sites (USERPOPINFO =1) yielded similar results as the analysis without this supplementary information. Testing with additional information regarding sample sites can bias the data towards grouping the individuals in accordance to the added sample information (Pritchard *et al.*, 2007). The best estimated K was still 2 when running STRUCTURE with additional population information, further supporting the $K = 2$ suggestion.

The LnP(D) results from STRUCTURE output files (with additional sample site information) were incorporated in Distruct (Rosenberg, 2004) to visualize the results for all tests (Figure 5). For all the individual runs $K = 2$ was visualised, even if $K = 1$ produced the best result for both individual rivers. One assumed population would be visualised as one colour (Q = 1 for all individuals), therefore $K = 2$ was depictured. From Figure 5b and 5c it can be demonstrated that both rivers shows a tendency of being grouped as one population when K = 2 demonstrating equal possibilities of belonging to both groups. Some tendency of further substructuring can be seen in Bellbekken as B3 shows some signs of being different from B1 and B2 (Figure 5c). (Appendix 6 shows comparable outputs from STRUCTURE incorporated in DISTRUCT for all six stations for $K = 3$ and $K = 4$, clearly indicating that $K = 2$ is the best K describing the data).

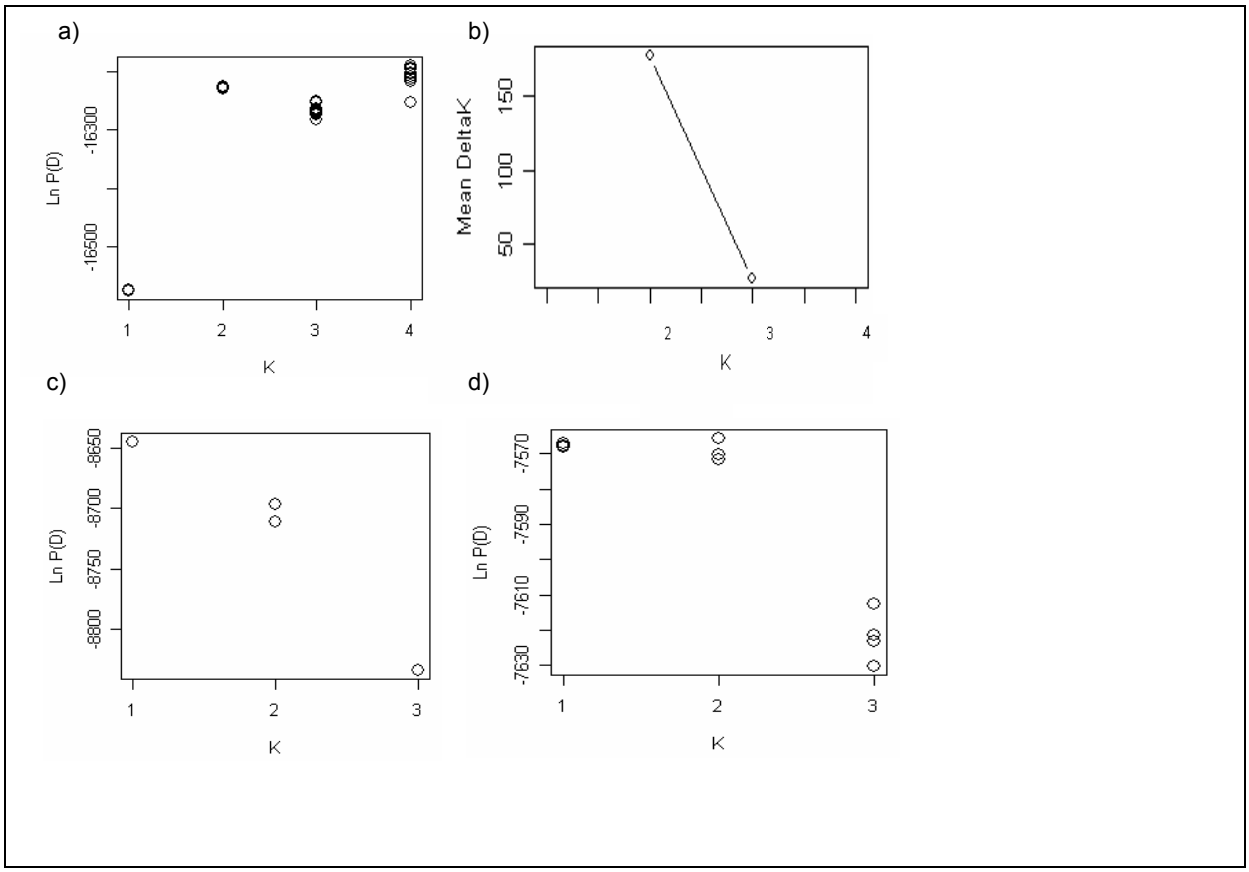


Figure 4. Results from STRUCTURE. Figure a) is showing the LnP(D) values plotted for each assumed K (1-4) for all the sample sites b) is showing ΔK values (Evanno *et al.*, 2005) for all the sample sites. c) is showing the estimated LnP(D) values for Julussa and d) is showing the estimated LnP(D) values for Bellbekken.

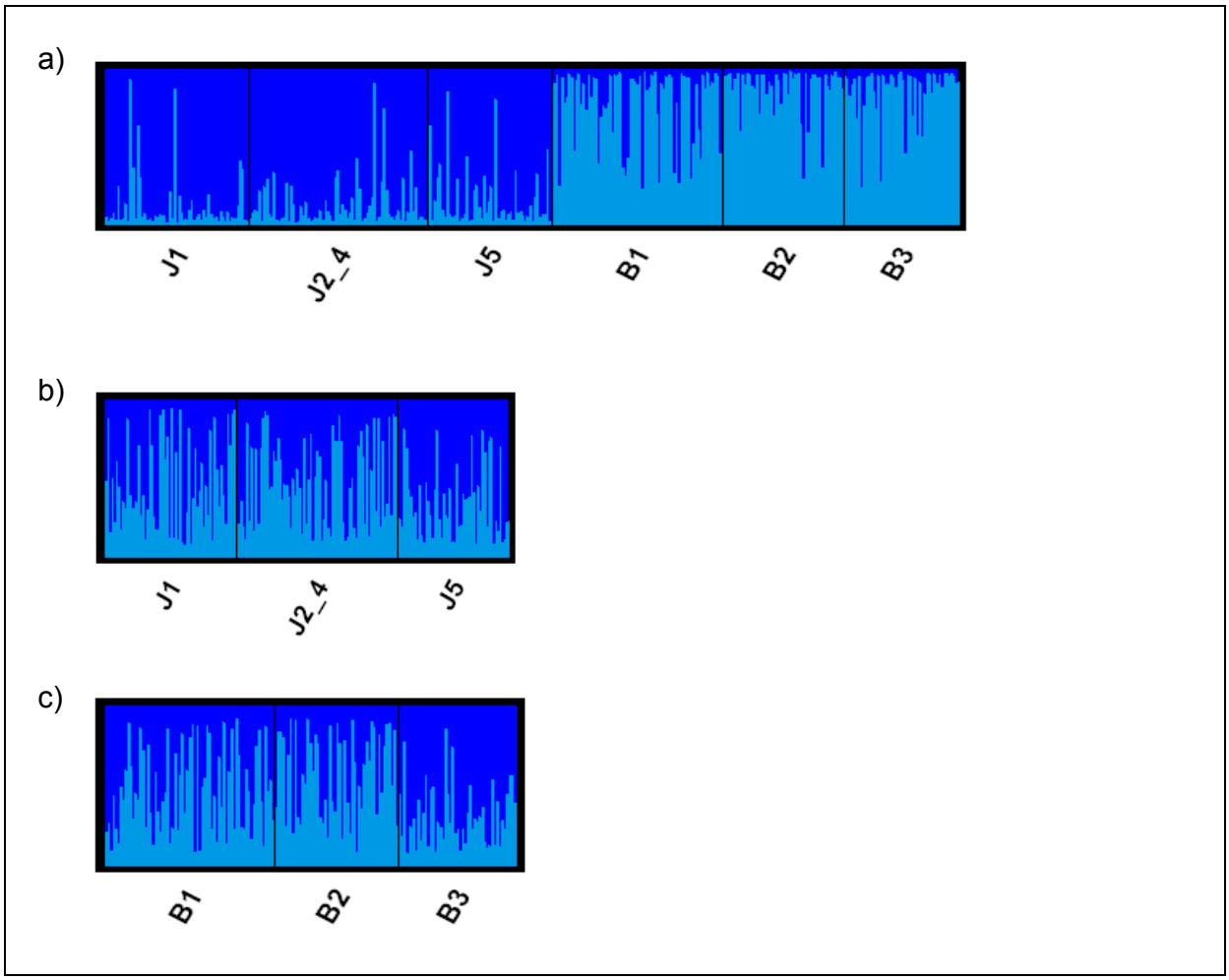


Figure 5. Summary plot of estimates of Q; a quantification of how likely each individual is belonging to each group (*K*) under consideration. a) represent the whole dataset, b) represent Julussa, c) represent Bellbekken. The figures express each individual by a thin vertical line. The vertical line is broken into *K* (2) coloured segments were the length of each colour segment represent the individual's membership in each of the clusters (Q). Sample site labels are shown at the bottom, sampling site information is shown in Table 1.

Discussion

This study yielded several interesting results. Firstly, it was clearly a substructuring between the two rivers, as could be seen from the F_{ST} values, the assignment test and the population structure test. Secondly, the isolation by distance test showed a positive relationship, marginally significant. Thirdly, further substructure within rivers was not evident, but B3 showed some signs of being differentiated, as this station had the highest self assignments and can to some extent be visually separated from B1 and B2 in the population structure test. No obvious substructure was detected within Julussa. All in all, these results suggest that the trout populations in the two different rivers are subdivided, not by distance itself, but through migration barriers.

Genetic differentiation between rivers

The studies of the genetic structuring of brown trout in Julussa and Bellbekken indicate a clear genetic difference between the two rivers. The F_{ST} values between sample sites in Bellbekken and Julussa ranged from 0.0216 (J1 - B1) to 0.0522 (J2_4 - B3) (Table 4). It has been suggested that F_{ST} values in the range of 0 - 0.05 indicate little genetic differentiation (Wright, 1978), but F_{ST} values it is not negligible at 0.005 or even less. This was already stressed by Wright (1978) and comes from the fact that the expectation of F_{ST} under complete differentiation not always will be one (because of polymorphism effects caused by mutations when migration is low) (Wright, 1978). The assignment test and population structure test also suggested a genetic subdivision between the two rivers. The partially Bayesian method (Rannala and Mountain, 1997) assigned a unnegligible fraction of individual fish back to the river they were obtained, when an average of 28 and 10% were assigned back to the river from which they were sampled, for 80 and 90% assignment thresholds, respectively (for both rivers combined) (Figure 3). Both rivers had a large fraction of unassigned individuals that could represent un-sampled subpopulations. The fully Bayesian method (Pritchard *et al.*, 2000) also structured the two rivers into two subpopulations (Figure 5a). In regard of this, these rivers should be considered as separate breeding populations.

This finding is in accordance with other studies concerning trout populations separated by a waterfall (Bouza *et al.*, 1999; Carlsson *et al.*, 1999; Hindar *et al.*, 1991; Ryman, 1983; Taylor *et al.*, 2003) as this arrange fish into different gene pools over and under the migration barrier. Carlsson and Nilsson (2001) were able to show that differentiation of populations of

resident brown trout was determined by a combination of impassable waterfalls, subdivision into tributaries and distance. Fish can move down waterfalls (Carlsson *et al.*, 1999; Morán *et al.*, 1995), but upward movement is often more difficult or even impossible. Therefore, for fish to maintain in a river above an impossible migration barrier, genetic differences resulting from a strong selection pressure for maintenance above the waterfall is expected, holding back potential downstream movement (Johnsson, 1989; Jonsson, 1982; Olsen and Vøllestad, 2001b). Downstream movement is reported, both with genetics (Figure 3) (four individuals) and as marked trout from Bellbekken has been re-caught in Julussa (Olsen and Vøllestad, unpublished data). There are also indications of fish migrating from Julussa and up Bellbekken (Figure 3) (two individuals), but to have any lasting effect regarding gene flow, the migrants also have to reproduce successfully. One effectively reproducing immigrant will not be sufficient to maintain identical allele frequencies between populations, but will be sufficient to ensure that the same alleles are shared (Allendorf and Phelps, 1981). The high levels of genetic differentiation between Julussa and Bellbekken indicate low gene flow between the rivers when only a few immigrants per generation with reproductive success would be sufficient to counteract the observed differentiation between the rivers (Spieth, 1974).

A weak trend of isolation by distance was found in the present study. Although the amount of variation of genetic distances that was explained by geographic distances was low ($r^2 = 6.9\%$), this support the tendency for populations being further apart also being more differentiated. Further, there was also evidence for the incorrectly classified individuals to be assigned to one of the nearest sampled sites (Figure 3), and when gene flow is more common between neighbouring localizations than between remote sections this can indicate some isolation by distance structure. This is also shown in tagged experiments with salmon (*Salmo salar*) and trout, when strayers (individuals spawning in "wrong" river) are most likely to ascend a river in close proximity to the natal river (Svärdson and Fagerström, 1982). However, the results are weak, therefore suggesting that on the spatial scale of this study, postglacial population structure has been determined also by other factors and not only by dispersal from a single refuge along the continuous rivers.

The lack of significant isolation by distance is in agreement with other studies (Carlsson and Nilsson, 2001; Crozier and Ferguson, 1986; Ryman, 1983) when a positive relationship seldom is reported from stationary brown trout in the literature (but see Estoup et al 1998, Carlsson et al 1999 and Carlsson and Nilsson 2000). The shortage of genetic differentiation concerning distance is suggested explained by the high genetic diversity within

the species, as a high amount of variation is distributed between populations regardless of distance between them (Ferguson, 1989). However, it is suggested by Hansen and Mensberg (1998) that the shortage of observations of isolation by distance in the literature is due to studies including migration barriers that is hindering gene flow, leaving the populations reproductively isolated where drift can lead to unidirectional changes of allele frequencies. This is supported as increasing genetic differentiation with distance has been found in sea trout (Hansen and Mensberg, 1998; Morán *et al.*, 1995) not separated by physical barriers. Several studies with isolated salmon populations find no significant pattern of isolation by distance (Meldgaard *et al.*, 2003), or the evidence for isolation by distance increases when populations above migration barriers where removed (Taylor *et al.*, 2003). Reproductive isolation could be part of the explanation of the insufficiency of significant isolation by distance in this study, when Bellbekken, with small geographical distances between sites express high genetic divergences (Table 4), probably due to drift, compared to Julussa, where longer distances between sites express low genetic differences (Table 4), probably due to higher migration. Carlsson and Nilsson (2000) point out that the lack of studies concerning several sampled localities within a single river could further explain the shortage of indications of isolation by distance

Genetic differentiation within rivers

Substructuring within rivers was not obvious, but Bellbekken showed significant differences of 0.5 - 2.13 % between sites separated by ~1 to 2.3 km (Table 4). Genetic differentiation within Bellbekken (Average F_{ST} = 0.0137) was similar to other studies (Carlsson and Nilsson 1999; within stream average F_{ST} = 0.0130). It would also be expected that Julussa should show some signs of differentiation, according to the longer distance between sample sites compared to Bellbekken (Figure 1), but no obvious differentiation could be detected in this stream, except for a significant differentiation between J1 and J5 of 1% (Table 4). That Bellbekken showed some signs of being further substructured was also indicated by the assignment test (Piry *et al.*, 2004) when B3 had the highest self-assignment (53 and 45% for 80 and 90% assignment thresholds, respectively) (Figure 3). The population structure test (Pritchard *et al.*, 2000) also indicated that B3 could be more substructured compared to B1 and B2 (Figure 5c).

That Bellbekken and Julussa showed little evidence for further substructuring is in accordance with other studies concerning within river differentiation (Hansen and Mensberg, 1998; Morán *et al.*, 1995). But, Bellbekken show some signs of further substructuring, as is

also revealed in other within river and lake studies not separated by migration barriers (Carlsson and Nilsson, 2000; Ferguson and Manson, 1981; Ryman *et al.*, 1979), showing that substructure without barriers is possible for this species, probably due to precise selection of spawning areas and homing by trout (Carlsson and Nilsson, 2001, Carlsson and Nilsson 2000). Traditionally, major genetic structure is considered to signal the potential existence of adaptive traits that enhance survival and reproduction in local environments (Carvalho, 1993). The most likely explanation for the higher microgeographic substructure observed in Bellbekken is not different genetic sources, but limited gene flow and genetic drift (Waples, 1998). The effective population size in Bellbekken is probably low, and with few individuals having reproductive success in each generation the populations will be subjects to genetic drift. When electrofishing B3 it was a marked reduction in fish abundance compared to B1 and B2, further supporting drift as the main cause for the substructuring observed. B3 also had lower heterozygosity than all other sample sites (Table 2), and it is known that reduced variation within populations tend to exaggerate measures of inter-population variation such as F_{ST} (Hedrick, 1999), again indicating drift, and not adaptation, as main cause for observed differentiation.

A higher migration rate is probably the reason for Julussa not being as substructured compared to Bellbekken. Even though brown trout are reported to be stationary (Bachman, 1984; Torgersen, 2006), the evidence for this is contradictive. Substatial movement in stream-dwelling brown trout populations are reported (Elliot, 1994; Gowan and Fausch, 1996a; Gowan and Fausch, 1996b; Gowan *et al.*, 1994; Heggenes *et al.*, 2007), where resident Salmonide seems to be far more mobile than previously reported, but with substantial individual variation depending on local ecological conditions and fish life-histories (Heggenes *et al.*, 2006). A high gene flow precludes local adaptation as it increases the effective population size and opposes random genetic drift. Julussa is a river of second order and has many small tributaries inhabited by brown trout draining into the river. This gives Julussa a higher potential for incoming fish; thereby a higher potential source of gene flow. At the same time, it could be that fish are migrating out to Julussa from smaller rivers to feed, for then to migrate up the same river (as they where born) when they are mature, for spawning (Elliot, 1994). The samples collected in Julussa could then be composed of several populations grouped together, which could further explain why Julussa seems so unstructured. Significant linkage disequilibrium (Knutsen *et al.*, 2001, and references therein) and deviation from Hardy-Weinberg equilibrium is expected if the sample sites consists of several populations; none which are reported for Julussa populations, thereby weakening this explanation.

In this study, only three localizations were sampled in each of the continuous rivers. This is not enough to draw conclusions about isolation by distance within river, but small trends can be viewed in both rivers (Figure 2). If more sample sites, at greater geographical distances, had been included in each river, the trend could have increased, as shown in Carlsson and Nilsson (2000).

Samples and tests

It is important to demonstrate temporal stability within a sampled site, as no further conclusions about the genetic structure observed could be considered valid without this stability. Errors, as artefacts of sampling just one cohort or a few families (Allendorf and Phelps, 1981; Hansen *et al.*, 1997), or variation in drift between years, can lead to samples collected at the same site being genetically differentiated between years; thereby overestimating spatial differentiation. Sampling over a restricted spatial scale (150-300 m) during a limited amount of time could have lead to a collection of related individuals (Hansen *et al.*, 1997), and is believed to be further biased if the sample contains a large proportion of fry (Allendorf and Phelps, 1981) as fry often have limited movement the first year and often are from the same redd (Elliot, 1994). In B1, when this station was sampled all the three sessions, and all the fish were released after sampling, the same fish could have been caught several times. As B1 have been part of an earlier capture - mark - recapture study, this could be ruled out since older fish had been individually marked, and the young of the year could be identified due to length. All the sampled localities include brown trout of different length (thereby different age) and should therefore not be heavily affected by family effects. The samples include young of the year, but not in a large proportion. Family effects could however affect the samples if some individuals have been spawning over several seasons; then having offspring in several generations. The differentiation between populations here was strong and no other evidence (no samples reported deviations from Hardy-Weinberg equilibrium due to homozygous excess) is given from the data that family effects could have affected the outcome of the data; therefore it is believed that similar result would have been obtained regardless of sampling procedure. In conclusion, this study reports temporal stability, whose long term stability can be confirmed only with studies covering longer time series.

The assignment test using the partially Bayesian method (Rannala and Mountain, 1997) has been shown to be the most accurate of the frequency assignment approaches (Cornuet *et al.*, 1999). At the same time, it has been shown that the fully Bayesian method

(Pritchard *et al.*, 2000) works just as well, or at some times, more accurate (Eldridge *et al.*, 2001).

In the partially Bayesian method (the assignment/exclusion test), as warned by Cornet et al (1999) two different types of errors did affect the assignment of individuals. Both error A and error E was evident in the dataset when running the test. Type A errors (when the correct population is not listed) occurred, where the solution was to assign the individuals to the river where assignment probabilities was the highest. Some exclusion to error E (the correct population is listed in addition to other possible populations of origin) was done when applying different stringencies to the test (Berry *et al.*, 2004) (80 and 90% assignment thresholds). Higher stringency resulted in fewer individuals assigned, both to specific sample sites and to river. This indicates that for some individuals, several sample sites could be regarded as potential populations of origin, and assignment was done to river instead of stations, when it was clear that the individual came from the specific river, just not which specific sample site it belonged to the most.

In contrast to the partially Bayesian assignment method, the fully Bayesian structure method assumes that the true candidate population is included in the analysis, and the posterior probabilities that an individual originated from each of the candidate populations sum to one. This means that if the true population is not sampled, but the individual had an initial probability of being assigned to a population that is sampled, this sampled population will have a false high assignment for that individual. Say, in reality there are three populations; A, B and C where the individual in focus have a true assignment probability of 0.99899, 0.001 and 0.00001 respectively. If population A not was sampled, this would lead the programme to estimate an assignment to population B with Q = 0.999 because the method compare relative probabilities. This could probably have biased the results as type A and E errors were reported from the assignment method as described above. Both tests should be used when the true population of origin might not have been sampled in the data set (Manel *et al.*, 2002), as parallel results (as obtained here) strengthen the results. It is believed that the accuracy would improve if more loci could be included (Berry *et al.*, 2004; Cornuet *et al.*, 1999).

Conclusion

The main result of the present work is the observation of two genetic divergent groups of brown trout populations in two rivers. This genetic structure is probably maintained by a small waterfall, acting as a migration barrier. Further, a weak support for isolation by distance was found. There was also a tendency for substructure within one of the rivers (Bellbekken) probably as a result of drift and limited gene flow.

References

Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research, **25**, 4692-4693.

Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. Canadian Journal of Fisheries and Aquatic Sciences, **38**, 1507-1514.

Bachman RA (1984) Foraging Behaviour of Free-Ranging Wild and Hatchery Brown Trout in a Stream. Transactions of the American Fisheries Society, **113**, 1-32.

Behnke RJ (1986) Brown trout. Trout, **27**, 42-47.

Berry O, Tocher MD, Sarre SD (2004) Can assignment tests measure dispersal? Molecular Ecology, **13**, 551-561.

Bohlin T, Hamrin S, Heggberget TG, Rasmussen G, Saltveit SJ (1989) Electrofishing - theory and practice with special emphasis on Salmonids. Hydrobiologia, **173**, 9-43.

Bouza C, Arias J, Castro J, Sánchez L, Martínez P (1999) Genetic structure of brown trout, *Salmo trutta* L., at the southern limit of the distribution range of the anadromous form. Molecular Ecology, **8**, 1991-2001.

Carlsson J, Nilsson J (2000) Population genetic structure of brown trout (*Salmo trutta* L.) within a northern boreal forest stream. Hereditas, **132**, 173-181.

Carlsson J, Nilsson J (2001) Effects of geomorphological structures on genetic differentiation among brown trout populations in a Northern Boreal river drainage. Transactions of the American Fisheries Society, **130**, 36-45.

Carlsson J, Olsen KH, Nilsson J, Øverli Ø, Stabell OB (1999) Microsatellites reveal fine-scale genetic structure in stream-living brown trout. Journal of Fish Biology, **55**, 1290-1303.

Carvalho GR (1993) Evolutionary aspects of fish distribution: genetic variability and adaptation. Journal of Fish Biology, **43**, 53-73.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics, **153**, 1989-2000.

Crozier WW, Ferguson A (1986) Electrophoretic examination of the population-structure of Brown Trout, *Salmo-Trutta* L, from the Lough Neagh catchment, Northern-Ireland. Journal of Fish Biology, **28**, 459-477.

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric Tandem Repeats. American Journal of Human Genetics, **49**, 746-756.

Ehrich D (2006) AFLPdat: a collection of R functions for convenient handling of AFLP data. Molecular Ecology Notes, **6**, 603-604.

Eldridge MDB, Kinnear JE, Onus ML (2001) Source population of dispersing rock-wallabies (*Petrogale lateralis*) identified by assignment tests on multilocus genotypic data. Molecular Ecology, **10**, 2867-2876.

Elliot JM (1994) Quantitative ecology and the brown trout. Oxford University Press, Oxford, UK.

Estoup A, Gharbi K, SanCristobal M, et al. (1998a) Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. Canadian Journal of Fisheries and Aquatic Sciences, **55**, 715-725.

Estoup A, Rousset F, Michalakis Y, et al. (1998b) Comparative analysis of microsatellite and allozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). Molecular Ecology, **7**, 339-353.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, **14**, 2611-2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics, **164**, 1567-1587.

Ferguson A (1989) Genetic differences among Brown Trout, *Salmo-Trutta*, stocks and their importance for the conservation and management of the species. Freshwater Biology, **21**, 35-46.

Ferguson A, Manson FM (1981) Allozyme evidence for reproductively isolated sympatric populations of brown trout *Salmo trutta* L. in Lough Melvin, Ireland. Journal of Fish Biology, **18**, 629-642.

Goudet J (1995) FSTAT (version 1.2): a computer programme to calculate F-statistics. Journal of Heredity, **86**, 485-486.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3) Available at http://www.unil.ch/softwares/fstat.html.

Goudet J, Raymond M, deMeeus T, Rousset F (1996) Testing differentiation in diploid populations. Genetics, **144**, 1933-1940.

Gowan C, Fausch KD (1996a) Long-term demographic responses of trout populations to habitat manipulation in six Colorado streams. Ecological Applications, **6**, 931-946.

Gowan C, Fausch KD (1996b) Mobile brook trout in two high-elevation Colorado streams: Re-evaluating the concept of restricted movement. Canadian Journal of Fisheries and Aquatic Sciences, **53**, 1370-1381.

Gowan C, Young MK, Fausch KD, Riley SC (1994) Restricted movement in resident stream Salmonids - a paradigm lost. Canadian Journal of Fisheries and Aquatic Sciences, **51**, 2626-2637.

Griffiths SW, Armstrong JD (2002) Kin-biased territory overlap and food sharing among Atlantic salmon juveniles. Journal of Animal Ecology, **71**, 480-486.

Grimholt U, Drablos F, Jorgensen SM, Hoyheim B, Stet RJM (2002) The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics, **54**, 570-581.

Guo SW, Thompson EA (1992) Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. Biometrics, **48**, 361-372.

Hansen MM, Kenchington E, Nielsen EE (2001) Assigning individual fish to populations using microsatellite DNA markers. Fish and Fisheries, **2**, 93-112.

Hansen MM, Mensberg KLD (1998) Genetic differentiation and relationship between genetic and geographical distance in Danish sea trout (*Salmo trutta* L.) populations. Heredity, **81**, 493-504.

Hansen MM, Nielsen EE, Mensberg KLD (1997) The problem of sampling families rather than populations: relatedness among individuals in samples of juvenile brown trout (*Salmo trutta* L.) Molecular Ecology, **6**, 469-474.

Hedrick PW (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. Evolution, **53**, 313-318.

Heggenes J, Omholt PK, Kristiansen JR, et al. (2007) Movements by wild brown trout in a boreal river: response to habitat and flow contrasts. Fisheries Management and Ecology, **14**, 333-342.

Heggenes J, Qvenild T, Stamford MD, Taylor EB (2006) Genetic structure in relation to movements in wild European grayling (*Thymallus thymallus*) in three Norwegian rivers. Canadian Journal of Fisheries and Aquatic Sciences, **63**, 1309-1319.

Hindar K, Johnsson B, Ryman N, Ståhl G (1991) Genetic relationships among landlocked, resident, and anadromous Brown trout, *Salmo trutta* L. Heredity, **66**, 83-91.
Johnsson B (1989) Life history and habitat use of Norwegian brown trout (*Samo trutta*). Freshwater Biology, **21**, 71-86.
Jonsson B (1982) Diadromus and resident trout *Salmo trutta*: is their difference due to genetics? Oikos, **38**, 297-300.
King TL, Eackles MS, Letcher BH (2005) Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. Molecular Ecology Notes, **5**, 130-132.
Knutsen H, Knutsen JA, Jorde PE (2001) Genetic evidence for mixed origin of recolonized sea trout populations. Heredity, **87**, 207-214.
Manel S, Berthier P, Luikart G (2002) Detecting wildlife poaching: Identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. Conservation Biology, **16**, 650-659.
Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Research, **27**, 209-220
Meldgaard T, Nielsen EE, Loeschcke V (2003) Fragmentation by weirs in a riverine system: A study of genetic variation in time and space among populations of European grayling (*Thymallus thymallus*) in a Danish river system. Conservation Genetics, **4**, 735-747.
Morán P, Pendás AM, García-Vázquez E, Izquierdo JI, Lobón-Cerviá J (1995) Estimates of gene flow among neighboring populations of brown trout. Journal of Fish Biology, **46**, 593-602.
Olsen EM, Vøllestad LA (2001a) Estimates of survival of stream-dwelling brown trout using mark-recaptures. Journal of Fish Biology, **59**, 1622-1637.
Olsen EM, Vøllestad LA (2001b) Within-stream variation in early life-history traits in brown trout. Journal of Fish Biology, **59**, 1579-1588.
Oreilly PT, Hamilton LC, McConnell SK, Wright JM (1996) Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. Canadian Journal of Fisheries and Aquatic Sciences, **53**, 2292-2298.
Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population-structure in Canadian polar bears. Molecular Ecology, **4**, 347-354.
Paetkau D, Shields GF, Strobeck C (1998) Gene flow between insular, coastal and interior populations of brown bears in Alaska. Molecular Ecology, **7**, 1283-1292.
Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Molecular Ecology, **13**, 55-65.
Paterson S, Piertney SB, Knox D, Gilbey J, Verspoor E (2004) Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. Molecular Ecology Notes, **4**, 160-162.
Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, **6**, 288-295.
Piry S, Alapetite A, Cornuet JM, et al. (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity, **95**, 536-539.
Poteaux C, Bonhomme F, Berrebi P (1999) Microsatellite polymorphism and genetic impact of restocking in Mediterranean brown trout (*Salmo trutta* L.). Heredity, **82**, 645-653.
Presa P, Guyomard R (1996) Conservation of microsatellites in three species of salmonids. Journal of Fish Biology, **49**, 1326-1329.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics, **155**, 945-959.
Pritchard JK, Wen X, Falush D (2007) Documentation for STRUCTURE software: Version 2.2. Avialeble from http://pritch.bsd.uchicago.edu/software
Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the United States of America, **94**, 9197-9201.
Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism Journal of Heredity, **86**, 248-249.
Rice WR (1989) Analysing tables of statistical tests. Evolution, **43**, 223-225.
Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes, **4**, 137-138.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics, **145**, 1219-1228.
Ryman N (1983) Patterns of distribution of biochemical genetic variation in salmonids: Differences between species. Aquaculture, **33**, 1-21.
Ryman N, Allendorf FW, Stahl G (1979) Reproductive isolation with little genetic divergence in Sympatric Populations of Brown Trout (*Salmo-Trutta*). Genetics, **92**, 247-262.
Skaala Ø (1992) Genetic population structure of Norwegina brown trout. Journal of Fish Biology, **41**, 631-646.
Slatkin M (1985) Gene flow in natural populations. Annual review of Ecology and Systematics, **16**, 393-430.
Spieth PT (1974) Gene flow and genetic differentiation. Genetics, **78**, 961-965.
Sugg DW, Chesser RK, Dobson FS, Hoogland JL (1996) Population genetics meets behavioral ecology. Trends in Ecology & Evolution, **11**, 338-342.
Svärdson G, Fagerström Å (1982) Adaptive differences in the long-distance migration of some trout (*Salmo trutta L*) stocks. Institute of freshwater research, Drotningsholm. **60**, 51-80.
Taylor EB (1991) A review of local adaptation in Salmonidae, with Particular Reference to Pacific and Atlantic salmon. Aquaculture, **98**, 185-207.
Taylor EB, Stamford MD, Baxter JS (2003) Population subdivision in westslope cutthroat trout (*Oncorhynchus clarki lewisi)* at the northern periphery of its range: evolutionary inferences and conservation implications. Molecular Ecology, **12**, 2609-2622.
Torgersen P (2006) Growth variation in brown trout *Salmo trutta* across small spatial scales: effects of density dependence and habitat quality? Master of Science Thesis, Department of Biology, University of Oslo.
Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes, **4**, 535-538.
Vasemagi A, Nilsson J, Primmer CR (2005) Seventy-five EST-linked Atlantic salmon (*Salmo salar* L.) microsatellite markers and their cross-amplification in five salmonid species. Molecular Ecology Notes, **5**, 282-288.
Waples RS (1989) A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics, **121**, 379-391.
Waples RS (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. Journal of Heredity, **89**, 438-450.
Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology, **15**, 1419-1439.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population-Structure. Evolution, **38**, 1358-1370.

Wright S (1931) Evolution in mendelian populations. Genetics, **16**, 97-159.

Wright S (1978) Evolution and the Genetics of Populations, Vol. IV: Variability within and among natural populations. University of Chicago Press, Chicago.

Zamudio KR, Wieczorek AM (2007) Fine-scale spatial genetic structure and dispersal among spotted salamander (*Ambystoma maculatum*) breeding populations. Molecular Ecology, **16**, 257-274.

Østbye K, Amundsen PA, Bernatchez L, et al. (2006) Parallel evolution of ecomorphological traits in the European whitefish *Coregonus lavaretus* (L.) species complex during postglacial times. Molecular Ecology, **15**, 3983-4001.

Appendices

Appendix 1: Microsatellite information
Microsatellite locus names, primer sequence (Primers 5' - 3'); F: forward, R: reverse, FAM (Blue), NED (yellow), PET (red), VIC (green) dyes, repeated motifs when given; dinucleotides (Di) or tetranucleoides (Tet) loci and producer; Applied biosystems (App.Bio) and DNA technologies AS (DNA tec.).

Locus name	Primers 5' - 3'	Repeated motif	Producer:
CA040261	**F: VIC**-CAGAGAGGAACCCACGTCAC	Di: $(GT)_{13}$	App.Bio
	R: GTTTAGGCATGTACGCATTTAGGC		DNA Tec.
CA053307	**F: FAM**-CAATGGGACCATCTCCCTAA	Di: $(TC)_{21}$	DNA Tec.
	R: GTTTCAGACCAATCAACCTGC		DNA Tec.
CA060177	**F: PET**-CGCTTCCTGGACAAAAATTA	Tet: $(TGAG)_{18}$	App.Bio
	R: GTTTGAGCACACCCATTCTCA		DNA Tec.
MST-85	**F: VIC**-GGAAGGAAGGGAGAAAGG T	Di: $(CT)_{22}$	App.Bio
	R: GGAAAATCAATACTAACAA		DNA Tec.
Ssa-85	**A**: AGGTGGGTCCTCCAAGCTAC	Di: $(GT)_{14}$	DNA Tec.
	B: PET-ACCCGCTCCTCACTTAATC		App.Bio
SSaD58	**F: FAM**-TAGAGTTTGTTCTCTGGCTTTG	Tet: TAGA	DNA Tec.
	R: AGACCCTAGGACTGGCTACTG		DNA Tec.
SsaD71	**F**: **PET**-AACGTGAAACATAAATCGATGG	Tet: TAGA	App.Bio
	R: TTA AGAATGGGTTGCCTATGAG		DNA Tec.
SsaD85	**F: FAM**-CTT TGGCTGTTTCAGGTATGAC	Tet:	DNA Tec.
	R: CACTGCTCTACAACAGAAGTCTC		DNA Tec.
SSaD157	**F: PET**-ATCGAAATGGAACTTTTGAATG	Tet: TAGA	App.Bio
	R: GCT TAGGGCTGAGAGAGGAATAC		DNA Tec.
SsaD170	**F: NED**- GGAGGCAGTTAAGAGAACAAAAG	Tet:	App.Bio
	R: TCACCTACCCTTCTCATTCAAG		DNA Tec.
SSaD190	**F: FAM**-GGCATTGGAGGTAAGGACAC	Tet: TAGA	DNA Tec.
	R: CCAGACCACTGAACTTCTCATC		DNA Tec.
SSaD237	**F: VIC**-CAATGATGGAGTGGGAATTATC	Tet: TAGA	App.Bio
	R: CCTCTATCCATACAACACATGC		DNA Tec.
SSsp2213	**F: VIC** ATGTGGAGGTCAACTAACCAGCGTG	Tet: $(GTTA)_{22}$	App.Bio
	R: CATCAATCACAGAGTGAGGCACTCG		DNA Tec.
STR2INRA	**F: FAM**-GGTGGCCTGGGTATAGCC	Di: $(CT)_4(TG)_{31}$	DNA Tec.
	R: GGTGTCGTTCAGCTGTAGCG		DNA Tec.
Strutta-12	**F. NED**-AGCTATTTCAGACATCACC	Di: $(GT)_{43}$	App.Bio
	R: AATCTCAAATCGATCAGAAG		DNA Tec.
TAP2B	**F: NED**-GCGGGACACCGTCAGGGCAGT	Di:	App.Bio
	R: GTTTCCTGATATTGTCTGCCAG		DNA Tec.

Appendix 2: Individual locus setup for PCR amplification

F and R indicate forward and reverse primer sequence. The PCR program that the loci correspond to is named at the bottom for each loci (see Appendix 3). See the text for more information about the single ingredients.

MST-85	
DNA	1,0
10x	1,000
dNTP	0,300
F	0,400
R	0,400
mq H_2O	6,850
Taq	0,050
Program	*1*

SSA-85	
DNA	1,0
10 X	1,000
dNTP	0,300
F	0,300
R	0,300
mq H_2O	7,070
Taq	0,030
Program	*2*

SSaD157	
DNA	1,0
10x	1,000
dNTP	0,300
F	0,400
R	0,400
mq H_2O	6,850
Taq	0,050
Program	*3*

SSaD190	
DNA	1,0
10x	1,000
dNTP	0,300
F	0,300
R	0,300
mq H_2O	7,070
Taq	0,030
Program	*4*

STR-2INRA	
DNA	1,0
10x	1,000
dNTP	0,300
F	0,500
R	0,500
mq H_2O	6,650
Taq	0,050
Program	*5*

Stutta-12	
DNA	1,0
10x	1,000
dNTP	0,300
F	0,500
R	0,500
mq H_2O	6,650
Taq	0,050
Program	*6*

CA053307	
DNA	1,0
10x	1,000
dNTP	0,300
F	0,400
R	0,400
mq H_2O	6,850
Taq	0,050
Program	*7*

CA040261	
DNA	1,0
10x (no MgCl)	1,000
MgCl2 (25mM)	0,480
dNTP	0,300
F	0,250
R	0,250
mq H_2O	6,670
Taq	0,050
Program	*7*

SSsp 2213	
DNA	1,0
10x no Mg	1,000
MgCl2 (25 mM)	1,000
dNTP	0,300
F	0,300
R	0,300
mq H_2O	6,070
Taq	0,030
Program	*8*

SSaD237	
DNA	1,0
10x (no MgCl)	1,000
MgCl2 (25mM)	0,720
dNTP	0,300
F	0,400
R	0,400
mq H_2O	6,130
Taq	0,050
Program	*9*

SSaD58	
DNA	1,0
10x	1,000
dNTP	0,300
F	0,400
R	0,400
mq H_2O	6,870
Taq	0,030
Program	*9*

Triplex:

SsaD71 + SsaD85 + SsaD170	
DNA	1,5
10 X	1,500
dNTP	0,450
SsaD71-F	0,400
SsaD71-R	0,400
SsaD85-F	0,400
SsaD85-R	0,400
SsaD170-F	0,300
SsaD170-R	0,300
mq H_2O	9,300
Taq	0,050
Program	*9*

Duplex:

CA060177 + TAP2B	
DNA	1,0
10x	1,000
dNTP	0,300
CA050177-F	0,400
CA050177-R	0,400
TAP2B-F	0,200
TAP2B-R	0,200
mq H_2O	6,450
Taq	0,050
Program	*10*

Appendix 3: PCR program setup.
For all the 16 loci (Appendix 1, Appendix 2), 10 different setups were made. Degrees of Celsius (ºC) with corresponding time (in seconds and minutes) and intervals shown for the different programs. See the text for more details.

Program 1

94 ºC	3 min	
92 ºC	45 sec	x 30
55 ºC	45 sec	
72 ºC	1 min	
72 ºC	10 min	
4 ºC	∞	

Program 2

95 °C	2 min	
95 °C	30 sec	x 35
56 °C	45 sec	
72 °C	45 sec	
72 °C	5 min	
4 °C	∞	

Program 3

94 ºC	5 min	
94 ºC	45 sec	x 34
56 ºC	45 sec	
68 ºC	2 min	
68 ºC	10 min	
4 ºC	∞	

Program 4

94 ºC	5 min	
94 ºC	45 sec	x 34
66 ºC	45 sec	
68 ºC	2 min	
68 ºC	10 min	
4 ºC	∞	

Program 5

94 ºC	5 min	
94 ºC	30 sec	x 35
58 ºC	30 sec	
72 ºC	30 sec	
72 ºC	10 min	
4 ºC	∞	

Program 6

94 ºC	3 min	
92 ºC	45 sec	x 30
57 ºC	45 sec	
72 ºC	1 min	
72 ºC	5 min	
4 ºC	∞	

Program 7

94 ºC	2 min	
94 ºC	30 sec	x 35
60 ºC	30 sec	
72 ºC	30 sec	
72 ºC	10 min	
4 ºC	∞	

Program 8

95 °C	2 min	
95 °C	30 sec	x 35
58 °C	30 sec	
72 °C	1 min	
72 °C	5 min	
4 °C	∞	

Program 9

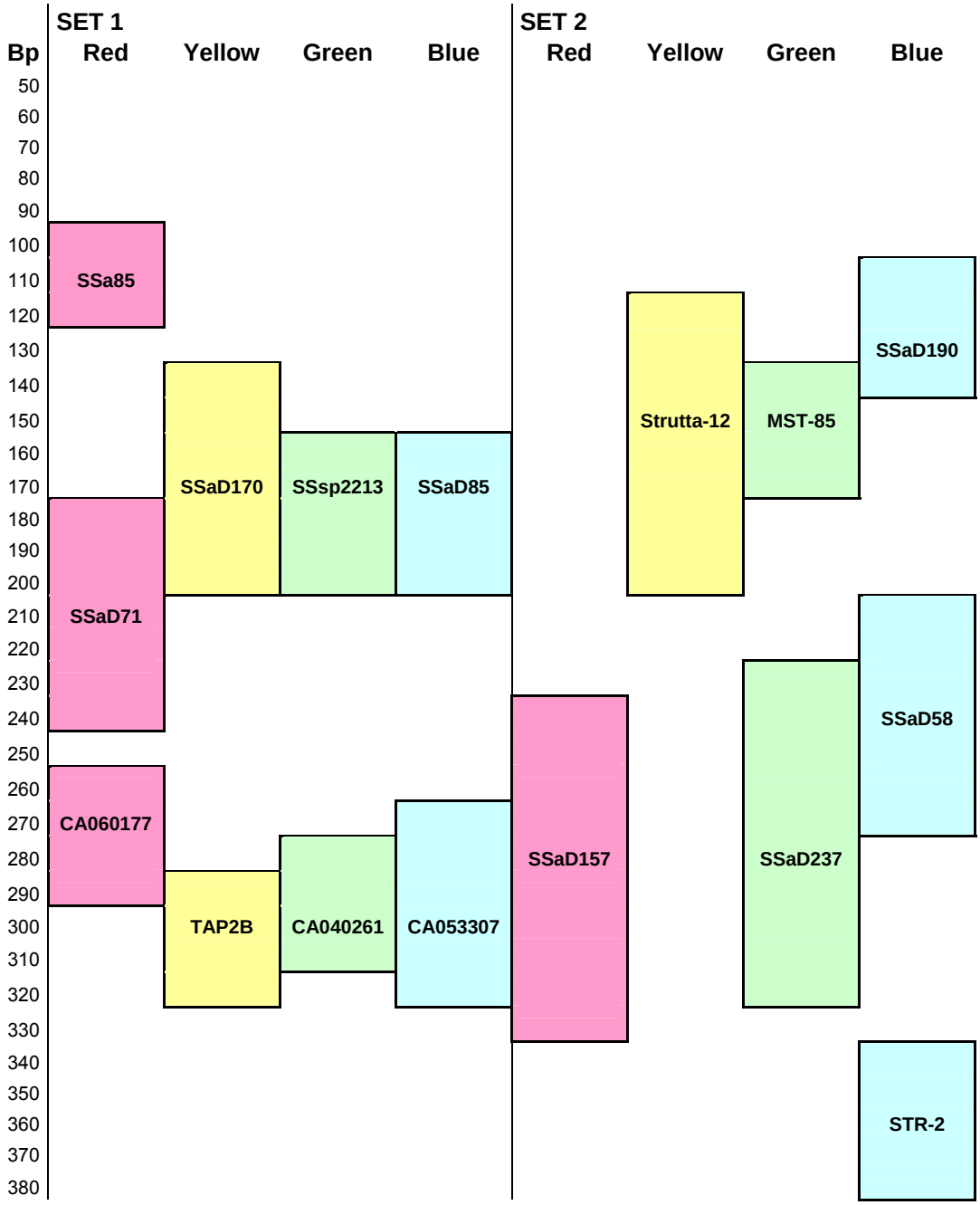
94 °C	5 min	
94 °C	45 sec	x 34
60 °C	45 sec	
68 °C	2 min	
68 °C	10 min	
4 °C	∞	

Program 10

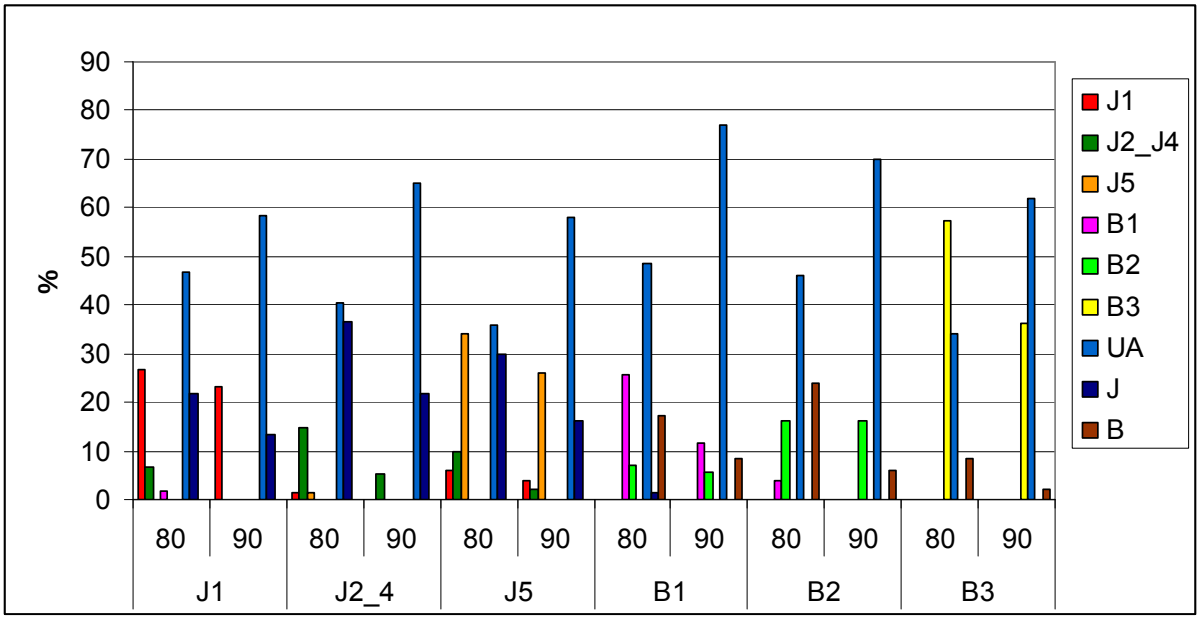
94 ºC	2 min	
94 ºC	30 sec	x20
* 60 ºC	30 sec	
72 ºC	30 sec	
94 ºC	30 sec	x15
60 ºC	30 sec	
72 ºC	30 sec	
72 ºC	5 min	
4 ºC	∞	

* Touchdown

Appendix 4. Genotyping setup.
The base pare length (Bp) and colourings of the different loci is shown. The different length and colourings made it possible to make two different sets of the 16 microsatellites (Appendix 1) in order to genotype the locus individually on Genemapper (Applied biosystems).



Appendix 5. Individual assignment test.
Assignment test with 10 000 simulated individuals. The results are given for tests with an assignment threshold with 80% (80) and 90% (90), values reported in percent. Unassigned individuals and individuals assigned Julussa (J) and Bellbekken (B), when individuals were assigned to multiple stations within the same rivers, are reported for each sample site and threshold. Sampling site information is shown in Table 1.



Appendix 6. Population structure test.
Summary plot of estimates of Q; a quantification of how likely each individual is belonging to each group (*K)* under consideration. a) represent the whole dataset where *K* = 3, b) represent the whole dataset where *K* = 4. The figures express each individual by a thin vertical line. The vertical line is broken into *K* coloured segments were the length of each colour segment represent the individual's membership in each of the clusters (Q). Sample site labels are shown at the bottom, sampling site information is shown in Table 1.